

A Short Introduction to Epidemiology

Second Edition

Neil Pearce

Occasional Report Series No 2

Centre for Public Health Research
Massey University Wellington Campus
Private Box 756
Wellington, New Zealand

Centre for Public Health Research
Massey University Wellington Campus
Private Box 756
Wellington, New Zealand
Phone: 64-4-3800-606
Fax: 64-4-3800-600
E-mail: cphr@massey.ac.nz
Website: <http://www.publichealth.ac.nz/>

Copies of this publication can be purchased in hard copy through our website (NZ\$36.744 incl GST), or downloaded for free in pdf form from the website.

2nd edition

February 2005

ISBN 0-476-01236-8

ISSN 1176-1237

To Irihapeti Ramsden

Preface

Who needs another introductory epidemiology text? Certainly, there are many introductory epidemiology books currently in print, and many of them are excellent. Nevertheless, there are four reasons why I believe that this new text is justified.

Firstly, it is much shorter than most introductory texts, many of which contain more material than is required for a short introductory course. This is a short introduction to epidemiology, and is not intended to be comprehensive.

Secondly, I have endeavoured to show clearly how the different basic epidemiologic methods “fit together” in a logical and systematic manner. For example, I attempt to show how the different possible study designs relate to each other, and how they are different approaches to a common task. Similarly, I attempt to show how the different study design issues (confounding and other types of bias) relate to each other, and how the principles and methods of data analysis are consistent across different study designs and data types.

Thirdly, in this context, rather than attempt a comprehensive review of available methods (e.g. multiple methods for estimating confidence intervals for the summary risk ratio), I have attempted to select only one standard method for each application, which is reasonably robust and accurate, and which is consistent and coherent with the other methods presented in the text.

Finally, the field of epidemiology is changing rapidly, not only with regards to its basic methods, but also with regards to the hypotheses which these methods

are used to investigate. In particular, in recent years there has been a revival in public health applications of epidemiology, not only at the national level, but also at the international level, as epidemiologists tackle global problems such as climate change. This text does not attempt to review the more complex measures used to consider such issues. However, it does provide a coherent and systematic summary of the basic methods in the field, which can be used as a logical base for the teaching and development of research into these more complex issues.

Chapter 1 gives a brief introduction to the field, with an emphasis on the broad range of applications and situations in which epidemiologic methods have been used historically, and will continue to be used in the future.

Part 1 then addresses study design options. Chapter 2 discusses incidence studies (including cohort studies) and describes the basic study design and the basic effect measures (i.e. incidence rates and rate ratios). It then presents incidence case-control studies as a more efficient means of obtaining the same findings. Chapter 3 similarly discusses prevalence studies, and prevalence case-control studies. Chapter 4 then considers study designs incorporating other axes of classification, continuous outcome measures (e.g. blood pressure) such as cross-sectional studies and longitudinal studies, or more complex study designs such as ecologic and multi-level studies.

Part 2 then addresses study design issues. Chapter 5 discusses issues of study size and precision. Chapter 6 considers general issues of validity, namely selection bias, information bias, and confounding. Chapter 7 discusses effect modification.

Part 3 then discusses the practical issues of conducting a study. Chapter 8 addresses issues of measurement of exposure and disease. Chapters 9-11 then discuss the conduct of cohort studies, case-control studies and cross-sectional studies respectively.

Finally, Part 4 considers what happens after the data are collected, with chapter 12 addressing data analysis and chapter 13 the interpretation of the findings of epidemiologic studies.

I should stress that this book provides no more than a very preliminary introduction to the field. In doing so I have attempted to use a wide range of examples, which give some indication of the broad range

of situations in which epidemiologic methods can be used. However, there are undoubtedly many other types of epidemiologic hypotheses and epidemiologic studies which are not represented in this book. In particular, my focus is on the use of epidemiology in public health, particularly with regard to non-communicable disease, and I include few examples from clinical epidemiology or from communicable disease outbreak investigations. Nevertheless, I hope that the book will be of interest not only to epidemiologists, but also to others who have other training but are involved in epidemiologic research, including public health professionals, policy makers, and clinical researchers.

Neil Pearce

*Centre for Public Health Research
Massey University Wellington Campus
Private Box 756
Wellington, New Zealand*

Acknowledgements

During the writing of this text, my salary was funded by the Health Research Council of New Zealand. I wish to thank Sander Greenland and Jonny Myers for their comments on the draft manuscript. I also wish to thank Massey University for support for my research programme.

A Short Introduction to Epidemiology

Contents

1. Introduction	9		
– Germs and miasmas	10		
– Risk factor epidemiology	11		
– Epidemiology in the 21 st century	12		
PART 1: STUDY DESIGN OPTIONS			
2. Incidence studies	21		
– Incidence studies	22		
– Incidence case-control studies	28		
3. Prevalence studies	33		
– Prevalence studies	33		
– Prevalence case-control studies	38		
4. More complex study designs	41		
– Other axes of classification	41		
– Continuous outcome measures	42		
– Ecologic and multilevel studies	47		
PART 2: STUDY DESIGN ISSUES			
5. Precision	59		
– Basic statistics	60		
– Study size and power	61		
6. Validity	67		
– Confounding	67		
– Selection bias	73		
– Information bias	74		
7. Effect modification	83		
– Concepts of interaction	83		
– Additive and multiplicative models	88		
– Joint effects	89		
PART 3: CONDUCTING A STUDY			
8. Measurement of exposure and health status	95		
– Exposure	95		
– Health status	102		
–			
9. Cohort studies	109		
– Defining the source population and risk period	109		
– Measuring exposure	112		
– Follow-up	113		
10. Case-control studies	117		
– Defining the source population and risk period	117		
– Selection of cases	118		
– Selection of controls	119		
– Measuring exposure	122		
11. Prevalence studies	125		
– Defining the source population	125		
– Measuring health status	126		
– Measuring exposure	128		
PART 4: ANALYSIS AND INTERPRETATION OF STUDIES			
12. Data analysis	133		
– Basic principles	133		
– Basic analyses	136		
– Controlling for confounding	140		
13. Interpretation	145		
– Appraisal of a single study	145		
– Appraisal of all of the available evidence	148		

CHAPTER 1. Introduction

(In: Pearce N. A Short Introduction to Epidemiology. 2nd ed. Wellington, CPHR, 2005)

Public health is primarily concerned with the prevention of disease in human population. It differs from clinical medicine both in its emphasis on prevention rather than treatment, and in its focus on populations rather than individual patients (table 1.1). Epidemiology is the branch of public health which attempts to discover the causes of disease in order to make disease prevention possible. Epidemiological methods can be used in other contexts (particularly in clinical research), but this short introductory text focuses on the use of epidemiology in public health, i.e. on its use as part of the wider process of discovering the causes of disease and preventing its occurrence in human populations.

In this context, epidemiology has been defined as (Last, 1988):

"the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to control of health problems"

This broad definition could in theory include a broad range of research methodologies including qualitative research and quantitative randomised controlled trials. Some epidemiologists

recognise the complementary nature of the former (McKinlay, 1993), and some texts include the latter in their definition of epidemiology. However, the key feature of epidemiological studies is that they are quantitative (rather than qualitative) observational (rather than experimental) studies of the determinants of disease in human populations (rather than individuals). This will be my focus here, while recognising the value, and complementary nature, of other research methodologies. The observational approach is a major strength of epidemiology as it enables a study to be conducted in a situation where a randomized trial would be unethical or impractical (because of the large numbers of subjects required). It is also the main limitation of epidemiological studies in that the lack of randomization means that the groups being compared may differ with respect to various causes of disease (other than the main exposure under investigation). Thus, epidemiological studies, in general, experience the same potential problems as randomized controlled trials, but may suffer additional problems of bias because exposure has not been randomly allocated and there may be differences in baseline disease risk between the populations being compared.

Table 1.1

The defining features of public health: populations and prevention

	Prevention	Treatment
Populations	Public health	Health systems research
Individuals	Primary health care/ Health education	Medicine (including primary health care)

1.1 Germs and Miasmas

Epidemiology is as old as public health itself, and it is not difficult to find epidemiological observations made by physicians dating back to Hippocrates who observed that:

“Whoever wishes to investigate medicine properly should proceed thus: in the first place to consider the seasons of the year, and what effects each of them produces... when one comes into a city in which he is a stranger, he should consider its situation, how it lies as to the winds and the rising of the sun...One should consider most attentively the waters which the inhabitants use...and the ground... and the mode in which the inhabitants live, and what are their pursuits, whether they are fond of drinking and eating to excess, and given to indolence, or are fond of exercise and labor”. (Hippocrates, 1938; quoted in Hennekens and Buring, 1987)

Many other examples of epidemiological reasoning were published through the

ages. However, epidemiology was founded as an independent discipline in a number of Western countries in parallel with the industrial revolution of the 19th century. In Anglophone countries it is considered to have been founded by the work of Chadwick, Engels, Snow and others who exposed the appalling social conditions during the industrial revolution, and the work of Farr and others who revealed major socioeconomic differences in disease in the 19th century. At that time, epidemiology was generally regarded as a branch of public health and focused on the causes and prevention of disease in populations, in comparison with the clinical sciences which were branches of medicine and focussed on disease pathology and treatment of disease in individuals. Thus, the emphasis was on the prevention of disease and the health needs of the population as a whole. In this context, the fundamental importance of population-level factors (the urban environment, housing, socioeconomic factors, etc) was clearly acknowledged (Terris, 1987).

Table 1.2

Deaths and death rates from cholera in London 1854 in households supplied by the Southwark and Vauxhall Water Company and by the Lambeth Water Company

	Houses	Cholera deaths	Deaths per 10,000 houses
Southwark and Vauxhall	40,046	1,263	315
Lambeth Company	26,107	98	37
Rest of London	256,423	1,422	59

Source: (Snow, 1936; quoted in Winkelstein, 1995)

Perhaps the most commonly quoted epidemiologic legend is that of Snow who studied the causes of cholera in London in the mid-19th century (Winkelstein, 1995). Snow was able to establish that the cholera death rate was much higher in areas supplied by the Southwark and Vauxhall Company which took water from the Thames downstream from London (i.e. after it had been contaminated with sewerage) than in areas supplied by the Lambeth Company which took water from upstream, with the death rates being intermediate in areas served by both companies. Subsequently, Snow (1936) studied the area supplied by both companies, and within this area walked the streets to determine for each house in which a cholera death had occurred, which company supplied the water. The death rate was almost ten times as high in houses supplied with water containing sewerage (table 1.2).

Although epidemiologists and other researchers continue to battle over Snow's legacy and its implications for epidemiology today (Cameron and Jones,

1983; Loomis and Wing, 1991; Samet, 2000; Vandenbroucke, 1994), it is clear that Snow was able to discover, and establish convincing proof for, the mode of transmission of cholera, and to take preventive action several decades before the biological basis of his observations was understood. Thus, it was not until several decades after the work of Snow that Pasteur and others established the role of the transmission of specific pathogens in what became known as the "infectious diseases", and it was another century, in most instances, before effective vaccines or antibiotic treatments became available. Nevertheless, a dramatic decline in mortality from these diseases occurred from the mid-nineteenth century long before the development of modern pharmaceuticals. This has been attributed to improvements in nutrition, sanitation, and general living conditions (McKeown, 1979) although it has been argued that specific public health interventions on factors such as urban congestion actually played the major role (Szreter, 1988).

1.2 Risk Factor Epidemiology

This decline in the importance of communicable disease was accompanied by an increase in morbidity and mortality from non-communicable diseases such as heart disease, cancer, diabetes, and respiratory disease. This led to major developments in the theory and practice of epidemiology, particularly in the second half of the 20th century. There has been a particular emphasis on aspects of individual lifestyle (diet, exercise, etc) and in the last decade the

human genome project has seen an accelerated interest in the role of genetic factors (Beaty and Khoury, 2000).

Thus, epidemiology became widely recognized with the establishment of the link between tobacco smoking as a cause of lung cancer in the early 1950's (Doll and Hill, 1950; Wynder and Graham, 1950), although this association had already been established in Germany in the 1930s (Schairer and Schöninger,

2001). Subsequent decades have seen major discoveries relating to other causes of chronic disease such as asbestos, ionizing radiation, viruses, diet, outdoor air pollution, indoor air pollution, water pollution, and genetic factors. These epidemiologic successes have in some cases led to successful preventive interventions without the need for major social or political change. For example, occupational carcinogens can, with some difficulty, be controlled through regulatory measures, and exposures to known occupational carcinogens have been reduced in industrialized countries in recent decades. Another example is the successful World Health Organisation (WHO) campaign against smallpox. More recently, some countries have passed legislation to restrict advertising of tobacco and smoking in public places and have adopted health promotion programmes aimed at changes in "lifestyle".

Individual lifestyle factors would ideally be investigated using a randomised controlled trial, but this is often unethical or impractical (e.g. tobacco smoking). Thus, it is necessary to do observational studies and epidemiology has made major contributions to the understanding of the role of individual lifestyle factors and health. Because such factors would ideally be investigated in randomised controlled trials, and in fact would be ideally suited to such trials if it were not

for the ethical and practical constraints, epidemiologic theory and practice has, quite appropriately, been based on the theory and practice of randomised trials. Thus, the aim of an epidemiologic study investigating the effect of a specific risk factor (e.g. smoking) on a particular disease (e.g. lung cancer) is intended to obtain the same findings that would have been obtained from a randomised controlled trial. Of course, an epidemiologic study will usually experience more problems of bias than a randomised controlled trial, but the randomised trial is the "gold standard".

This approach has led to major developments in epidemiologic theory (presented most elegantly and comprehensively in Rothman and Greenland, 1998). In particular, there have been major developments in the theory of cohort studies (which mimic a randomised trial, but without the randomisation) and case-control studies (which attempt to obtain the same findings as a full cohort study, but in a more efficient manner). It is these basic methods, which follow a randomised controlled trial "paradigm", which receive most of the attention in this short introductory text. However, while presenting these basic methods, it is important to also recognise their limitations, and to also consider different or more complex methods that may be more appropriate when epidemiology is used in the public health context.

1.3 Epidemiology in the 21st Century

In particular, in the last decade there has been increasing concern expressed about the limitations of the risk factor approach, and considerable debate about

the future direction of epidemiology (Saracci, 1999). In particular, it has been argued that there has been an overemphasis on aspects of individual

lifestyle, and little attention paid to the population-level determinants of health (Susser and Susser, 1996a, 1996b; Pearce, 1996; McMichael, 1999). Furthermore, the success of risk factor epidemiology has been more temporary and more limited than might have been expected. For example, the limited success of legislative measures in industrialised countries has led the tobacco industry to shift its promotional activities to developing countries so that more people are exposed to tobacco smoke than ever before (Barry, 1991; Tominaga, 1986). Similar shifts have occurred for some occupational carcinogens (Pearce et al, 1994). Thus, on a global basis the "achievement" of the public health movement has often been to move public health problems from rich countries to poor countries and from rich to poor populations within the industrialized countries.

It should be acknowledged that not all epidemiologists share these concerns (e.g. Savitz, 1994; Rothman et al, 1998; Poole and Rothman, 1998), and some have regarded these discussions as an attack on the field itself, rather than as an attempt to broaden its vision. Nevertheless, the debate has progressed and there is an increasing recognition of the importance of taking a more global approach to epidemiologic research and of the importance of maintaining an appropriate balance and interaction between macro-level (population), individual-level (e.g. lifestyle), and micro-level (e.g. genetic) research (Pearce, 2004).

There are three crucial concepts which have received increasing attention in this regard.

The Importance of Context

The first, and most important issue, is the need to consider the population context when conducting epidemiologic

studies. Even if one is focusing on individual "lifestyle" risk factors, there is good reason to conduct studies at the population level (Rose, 1992). Moreover, every population has its own history, culture, and economic and social divisions which influence how and why people are exposed to specific risk factors, and how they respond to such exposures. For example, New Zealand (Aotearoa) was colonised by Great Britain more than 150 years ago, resulting in major loss of life by the indigenous people (the Māori). It is commonly assumed that this loss of life occurred primarily due to the arrival of infectious diseases to which Māori had no natural immunity. However, a more careful analysis of the history of colonisation throughout the Pacific reveals that the indigenous people mainly suffered major mortality from imported infectious diseases when their land was taken (Kunitz, 1994), thus disrupting their economic base, food supply and social networks. This example is not merely of historical interest, since it these same infectious diseases that have returned in strength in Eastern Europe in the last decade, after lying dormant for nearly a century (Bobak and Marmot, 1996). Similarly, the effects of occupational carcinogens may be greater in developing countries where workers may be relatively young or may be affected by malnutrition or other diseases (Pearce et al, 1994).

These issues are likely to become more important because, not only is epidemiology changing, but the world that epidemiologists study is also rapidly changing. We are seeing the effects of economic globalization, structural adjustment (Pearce et al, 1994) and climate change (McMichael, 1993, 1995), and the last few decades have seen the occurrence of the "informational revolution" which is having effects as great as the previous agricultural and industrial revolutions (Castells, 1996).

In industrialized countries, this is likely to prolong life expectancy for some, but not all, sections of the population. In developing countries, the benefits have been even more mixed (Pearce et al, 1994), while the countries of Eastern Europe are experiencing the largest sudden drop in life expectancy that has been observed in peacetime in recorded human history (Boback and Marmot, 1996) with a major rise in alcoholism and "forgotten" diseases such as tuberculosis and cholera.

This increased interest in population-level determinants of health has been particularly marked by increased interest in techniques such as multilevel modelling which allow individual lifestyle risk factors to be considered "in context" and in parallel with macro-level determinants of health (Greenland, 2000). Such a shift in approach is important, not only because of the need to emphasize the role of diversity and local knowledge (Kunitz, 1994), but also because of the more general moves within science to consider macro-level systems and processes (Cohen and Stewart, 1994) rather than taking a solely reductionist approach (Pearce, 1996).

Problem-Based Epidemiology

A second issue is that a problem-based approach may be particularly valuable in encouraging epidemiologists to focus on the major public health problems and to take the population context into account (Pearce, 2001; Thacker and Buffington, 2001). A problem-based approach to teaching clinical medicine has been increasingly adopted in medical schools around the world. The value of this approach is that theories and methods are taught in the context of solving real-life problems. Starting with "the problem" at the population level provides a "reality check" on existing etiological

theories and identifies the major public health problems which new theories must be able to explain. A fruitful research process can then be generated with positive interaction between epidemiologists and other researchers. Studying real public health problems in their historical and social context does not exclude learning about sophisticated methods of study design and data analysis (in fact, it necessitates it), but it may help to ensure that the appropriate questions are asked (Pearce, 1999).

Appropriate Technology

A related issue is the need to use "appropriate technology" to address the most important public health research questions. In particular, as attention moves "upstream" to the population level (McKinlay, 1993) new methods will need to be developed (McMichael, 1995). One example of this, noted above, is the recent rise in interest in multilevel modelling (Blakely and Woodward, 2000; Pearce, 2000), although it is important to stress that it is an increase in "multilevel thinking" in the development of epidemiologic hypotheses and the design of studies that is required, rather than just the use of new statistical techniques of data analysis. The appropriateness of any research methodology depends on the phenomenon under study: its magnitude, the setting, the current state of theory and knowledge, the availability of valid measurement tools, and the proposed uses of the information to be gathered, as well as the community resources and skills available and the prevailing norms and values at the national, regional or local level (Pearce and McKinlay, 1998). Thus, there has been increased interest in the interface between epidemiology and social science (Krieger, 2000), and in the

development of theoretical and methodological frameworks appropriate for epidemiologic studies in developing countries (Barreto et al, 2001; Barreto, 2004; Loewenson, 2004), and in indigenous people in "Western " countries (Durie, 2004). As

noted above, this short introductory text focuses on the most basic epidemiologic methods, but I attempt to refer to more complex issues, and the potential use of more complex methods, where this is appropriate.

Summary

Public health is primarily concerned with the prevention of disease in human populations, and epidemiology is the branch of public health which attempts to discover the causes of disease in order to make disease prevention possible. It thus differs from clinical medicine both in its emphasis on prevention (rather than treatment) and in its focus on populations (rather than individual patients). Thus, the epidemiological approach to a particular disease is intended to identify high-risk subgroups within the population, to determine the causes of such excess risks, and to determine the effectiveness of subsequent preventive measures. Although the epidemiological approach has been used for more than a century for the study of communicable diseases, epidemiology has considerably grown in scope and sophistication in the last few decades as it has been increasingly

applied to the study of non-communicable diseases. At the beginning of the 21st century, the field of epidemiology is changing rapidly, not only with regards to its basic methods, but also with regards to the hypotheses which these methods are used to investigate. In particular, in recent years there has been a revival in public health applications of epidemiology, not only at the national level, but also at the international level, as epidemiologists tackle global problems such as climate change. This text does not attempt to review the more complex methods used to study such issues. However, it does provide a coherent and systematic summary of the basic methods in the field, which can be used as a logical base for the teaching and development of research into these more complex issues.

References

- Barreto ML (2004). The globalization of epidemiology: critical thoughts from Latin America. *Int J Epidemiol* 33: 1132-7.
- Barreto ML, Almeida-Filho N, Breihl J (2001). Epidemiology is more than discourse: critical thoughts from Latin America. *J Epidemiol Comm Health* 55: 158-9.
- Barry M (1991). The influence of the U.S. tobacco industry on the health, economy, and environment of developing countries. *New Engl J Med* 324: 917-20.
- Beaty TH, Khoury MJ (2000). Interface of genetics and epidemiology. *Epidemiologic Reviews* 22: 120-5.
- Blakeley T, Woodward AJ (2000). Ecological effects in multi-level studies. *J Epidemiol Comm Health* 54: 367-74.
- Bobak M, Marmot M (1996). East-West mortality divide and its potential explanations: proposed research agenda. *Br Med J* 312: 421-5.
- Cameron D, Jones IG (1983). John Snow, the Broad Street pump and modern epidemiology. *Int J Epidemiol* 12: 393-6.
- Castells M (1996). The information age: Economy, society and culture. Vol 1. The rise of the network society. Oxford: Blackwell.
- Cohen J, Stewart I (1994). The collapse of chaos: discovering simplicity in a complex world. London: Penguin.
- Doll R, Hill AB (1950). Smoking and carcinoma of the lung. *Br Med J* 2: 739-48.
- Durie M (2004). Understanding health and illness: research at the interface between science and indigenous knowledge. *Int J Epidemiol* 33: 1138-43.
- Greenland S (2000). Principles of multilevel modelling. *Int J Epidemiol* 29: 158-67.
- Hennekens CH, Buring JE (1987). *Epidemiology in medicine*. Boston: Little, Brown.
- Hippocrates (1938). On airs, waters and places. *Med Classics* 3: 19.
- Krieger N (2000). Epidemiology and social sciences: towards a critical reengagement in the 21st century. *Epidemiologic Reviews* 22: 155-63.
- Kunitz S (1994). Disease and social diversity. New York: Oxford University Press.
- Last JM (ed) (1988). *A dictionary of epidemiology*. New York: Oxford University Press.
- Loewenson R (2004). Epidemiology in the era of globalization: skills transfer or new skills? *Int J Epidemiol* 33: 1144-50.
- Loomis D, Wing S (1991). Is molecular epidemiology a germ theory for the end of the twentieth century? *Int J Epidemiol* 19: 1-3.
- McMichael AJ (1993). *Planetary overload: global environmental change and the health of the human species*. Cambridge: Cambridge University Press.
- McMichael AJ (1995). The health of persons, populations, and planets: epidemiology comes full circle. *Epidemiol* 6: 633-5.
- McMichael AJ (1999). *Prisoners of the proximate: loosening the constraints*

- on epidemiology in an age of change. *Am J Epidemiol* 149: 887-97.
- McKeown T (1979). *The role of medicine*. Princeton, NJ: Princeton University Press.
- McKinlay JB (1993). The promotion of health through planned sociopolitical change: challenges for research and policy. *Soc Sci Med* 36: 109-17.
- Pearce N (1996). Traditional epidemiology, modern epidemiology, and public health. *AJPH* 86: 678-83.
- Pearce N (1999). Epidemiology as a population science. *Int J Epidemiol* 28: S1015-8.
- Pearce N (2000). The ecologic fallacy strikes back. *J Epidemiol Comm Health* 54: 326-7.
- Pearce N (2001). The future of epidemiology: a problem-based approach using evidence-based methods. *Australasian Epidemiologist* 8.1: 3-7.
- Pearce N (2004). The globalization of epidemiology: introductory remarks. *Int J Epidemiol* 33: 1127-31.
- Pearce N, McKinlay J (1998). Back to the future in epidemiology and public health. *J Clin Epidemiol* 51: 643-6.
- Pearce NE, Matos E, Vainio H, Boffetta P, Kogevinas M (eds) (1994). *Occupational cancer in developing countries*. Lyon: IARC.
- Poole C, Rothman KJ (1998). Our conscientious objection to the epidemiology wars. *J Epidemiol Comm Health* 52: 613-4.
- Rose G. *The strategy of preventive medicine*. Oxford: Oxford University Press, 1992.
- Rothman KJ, Greenland S (1998). *Modern epidemiology*. 2nd ed. Philadelphia: Lippincott-Raven.
- Rothman KJ, Adami H-O, Trichopoulos (1998). Should the mission of epidemiology include the readication of poverty? *Lancet* 352: 810-3.
- Samet JM (2000). Epidemiology and policy: the pump handle meets the new millennium. *Epidemiologic Reviews* 22: 145-54.
- Saracci R (1999). Epidemiology in progress: thoughts, tensions and targets. *Int J Epidemiol* 28: S997-9.
- Savitz DA (1994). In defense of black box epidemiology. *Epidemiology* 5: 550-2.
- Schairer E, Schöninger E (2001). Lung cancer and tobacco consumption. *Int J Epidemiol* 30: 24-7.
- Snow J (1936). *On the mode of communication of cholera*. (Reprint). New York: The Commonwealth Fund, pp 11-39.
- Susser M, Susser E (1996a). Choosing a future for epidemiology: I. Eras and paradigms. *Am J Publ Health* 86: 668-73.
- Susser M, Susser E (1996b). Choosing a future for epidemiology: II. From black boxes to Chinese boxes. *Am J Publ Health* 86: 674-8.
- Szreter S (1988). The importance of social intervention in Briatrain's mortality decline c.1850-1914: a reinterpretation of the role of public health. *Soc Hist Med* 1: 1-37.
- Terris M (1987). Epidemiology and the public health movement. *J Publ Health Policy* 7: 315-29.
- Thacker SB, Buffington J (2001). Applied epidemiology for the 21st century. *Int J Epidemiol* 30: 320-5.
- Tominaga S (1986). Spread of smoking to the developing countries. In: Zaridze D, Peto R (eds). *Tobacco: a major international health hazard*. Lyon: IARC, pp 125-33.

Vandenbroucke JP (1994). New public health and old rhetoric. *Br Med J* 308: 994-5.

Winkelstein W (1995). A new perspective on John Snow's communicable disease theory. *Am J Epidemiol* 142: S3-9.

Wynder EL, Graham EA (1950). Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma. *J Am Statist Assoc* 45: 329-38.

Part I

Study Design Options

CHAPTER 2. Incidence Studies

(In: Pearce N. A Short Introduction to Epidemiology. 2nd ed. Wellington, CPHR, 2005)

In this chapter and the next one I review the possible study designs for the simple situation where individuals are exposed to a particular risk factor (e.g. a particular chemical) and when a dichotomous outcome is under study (e.g. being alive or dead, or having or not having a particular disease). Thus, the aim is to estimate the effect of a (dichotomous) exposure on the occurrence of a (dichotomous) disease outcome or health state.

It should first be emphasized that all epidemiologic studies are (or should be) based on a particular source population (also called the "study population" or "base population") followed over a particular risk period. Within this framework a fundamental distinction is between studies of disease incidence (i.e. the number of new cases of disease over time) and studies of disease prevalence (i.e. the number of people with the disease at a particular point in time). Studies involving dichotomous outcomes can then be classified according to two questions:

- a. Are we studying studying incidence or prevalence?;
- b. Is there sampling on the basis of outcome?

The responses to these two questions yield four basic types of epidemiologic studies (Morgenstern and Thomas, 1993; Pearce, 1998):

1. Incidence studies
2. Incidence case-control studies
3. Prevalence studies
4. Prevalence case-control studies

These four study types represent cells in a two-way cross-classification (table 2.1). Such studies may be conducted to describe the occurrence of disease (e.g. to estimate the burden of diabetes in the community by conducting a prevalence survey), or to estimate the effect of a particular exposure on disease (e.g. to estimate whether the incidence new cases of diabetes is greater in people with a high fat diet than in people with a low fat diet) in order to find out how we can prevent the disease occurring. In the latter situation we are comparing the occurrence of disease in an "exposed" group with that in a "non-exposed" group, and we are estimating the effect of exposure on the occurrence of the disease, while controlling for other known causes of the disease.

Table 2.1

The four basic study types in studies involving a dichotomous health outcome

		Sampling on outcome	
		No	Yes
Study outcome	Incidence	Incidence studies	Incidence case-control studies
	Prevalence	Prevalence studies	Prevalence case-control studies

Thus, we might conclude that “lung cancer is five times more common in asbestos workers than in other workers, even after we have controlled for differences in age, gender, and smoking”. In some instances we may have multiple categories of exposure (high, medium, low) or individual exposure “scores”, but we will start with the simple situation in which individuals are classified as “exposed” or “non-exposed”.

In this chapter I discuss incidence studies, and in the following chapter I

consider prevalence studies. In chapter 4, I then consider studies involving more complex measurements of health status (e.g. continuous lung function or blood pressure measurements) and more complex study designs (ecologic and multilevel studies). As noted in chapter 1, the latter situation is perhaps the norm, rather than the exception, when conducting studies in the public health context. However, for logical and practical reasons I will first address the simpler situation of a dichotomous exposure (in individuals) and a dichotomous health outcome measure.

2.1 Incidence Studies

The most comprehensive approach involves collecting data on the experience of the entire source population over the risk period in order to estimate disease incidence (the development of a disease for the first time) or mortality (i.e. death which is a particular type of incidence measure). Figure 2.1 shows the experience of a source population in which all persons are followed from a particular date. For simplicity, I will initially assume that the source population is confined to persons born in a particular year, i.e. a birth cohort. In the hypothetical study shown in figure 2.1, the outcome under study is the “event” of developing a particular disease. However, the concept of incidence applies equally to studies of other health events, such as hospitalisation or death. The key feature of incidence studies is that they involve an event (e.g. developing a disease for the first time) which occurs at a particular point in time, rather than a state (e.g. having a disease) which can exist over an extended period of time.

In the hypothetical study shown in figure 2.1, people enter the study when they are born, and some of them subsequently develop disease. Of these, some subsequently “lose” their disease (although they may “regain” it at a later date), and some have the condition all their lives; some persons die from the disease under study, but most eventually die from another cause. However, the information is “censored” since the study cannot last indefinitely; i.e. follow-up stops by a particular age, at which time some members of the study population have died, and some have been lost to follow-up for other reasons (e.g. emigration). For example, several people in figure 2.1 were “censored” before follow-up finished, either because they died of the disease we were studying (if we were studying the incidence of disease, rather than deaths, they would be “censored” as soon as they developed the disease), they died of something else, or because they were “lost to follow-up”. Each person only contributes “person-time” to the study until they are

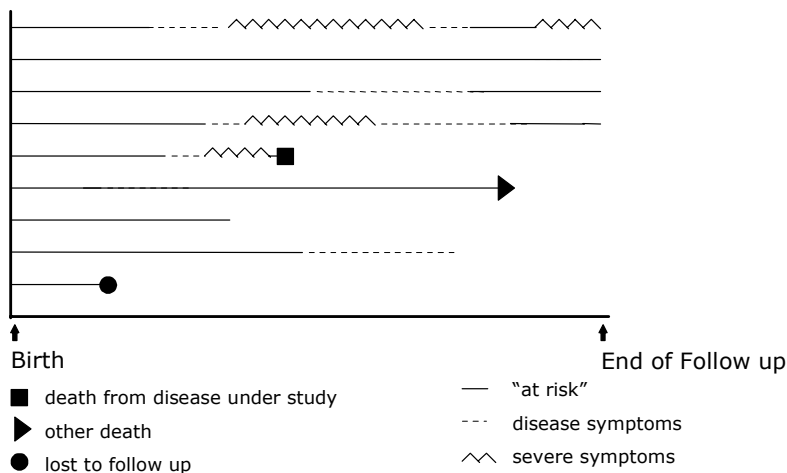
censored, and after that we stop counting them. This approach is followed because we may not get a fair comparison between the "exposed" and the "non-exposed" groups" if they have been followed for different lengths of time, e.g. if one group has many more people lost to follow-up than the other group.

However, the person-time approach would be necessary even if no-one was lost to follow up and both groups were followed for the same length of time. For example, consider a cohort study of 1,000 exposed and 1,000 non-exposed people in which no-one was lost to follow-up and everyone was followed until they died. Assume also that the exposure causes some deaths so the exposed group, on the average, died at a

younger age than the non-exposed group. If we only calculated the percentage of people who died, then it would be 100% in both groups, and we would see no difference. However, if we take into account the person-time contributed by each group, then it becomes clear that both groups had the same number of deaths (1,000), but that in the exposed group these deaths occurred earlier and the person-time contributed was therefore lower. Thus, the average age at death would be lower in the exposed group; to say the same thing another way, the death rate (deaths divided by person-years) would be higher. To see this, we need to consider not only how many people were in each group, but how much person-time they contributed, i.e. how long they were followed for.

Figure 2.1

Occurrence of disease in a hypothetical population followed from birth



Example 2.1

Martinez et al (1995) studied 1246 newborns in the Tucson, Arizona area enrolled between May 1980 and October 1984. Parents were contacted shortly after the children were born, and completed a questionnaire about their history or respiratory illness, smoking habits, and education. Further parental questionnaires

were completed during the child's second year of life and again at six years. At the age of six years, 51.5% of the children had never wheezed, 19.9% had had at least one lower respiratory tract illness with wheezing during the first three years of life but had no wheezing at six years, 15.0% had no wheezing before the age

of three years but had wheezing at six years, and 13.7% had wheezing both before three years of age and at six years. The authors concluded that the majority of infants with wheezing have transient conditions and do not have increased risks of asthma or allergies later in life.

In some circumstances, a study might be conducted to study the "natural history" of a disease (e.g. diabetes). In such "clinical epidemiology" studies, the population (denominator) under study comprises people who already have a particular disease or condition, and the goal is to ascertain which factors affect the disease prognosis. More typically, one might be interested in a particular hypothesis about developing disease, such as "a high cholesterol diet increases the risk of developing ischaemic heart disease". In this situation, the population under study comprises healthy individuals and we are interested in factors that determine who develops the disease under study (and who doesn't). The data generated by such an incidence study involve comparing "exposed" and "non-exposed" groups and are similar to that generated by a randomised controlled trial, except that dietary "exposure" has not been randomly allocated.

Incidence studies ideally measure exposures, confounders and outcome times on all population members. When the source population has been

formally defined and enumerated (e.g. a group of workers exposed to a particular chemical) then the study may be termed a cohort study or follow-up study (Rothman and Greenland, 1998) and the former terminology will be used here. Incidence studies also include studies where the source population has been defined but a cohort has not been formally enumerated by the investigator. Perhaps the most common examples are descriptive studies, e.g. of national death rates. In fact, as Rothman and Greenland (1998) note, no qualitative distinction distinguishes "descriptive" variables from the variables that are studied in "analytic" studies of risk factors. Thus, the distinction between "descriptive" incidence studies and "analytic" incidence studies is at best only a distinction based on data source (e.g. obtaining information from routine records rather than collecting the information specifically for the study).

Similarly, there is no fundamental distinction between incidence studies based on a broad population (e.g. all workers at a particular factory, or all

persons living in a particular geographical area) and incidence studies involving sampling on the basis of exposure, since the latter procedure merely redefines the source population (cohort) (Miettinen, 1985).

Measures of Disease Occurrence

I will briefly review the basic measures of disease occurrence that are used in incidence studies, using the notation depicted in table 2.2 which shows the findings of a hypothetical incidence study of 20,000 persons followed for 10 years (statistical analyses using these measures are discussed further in chapter 12).

Three measures of disease incidence are commonly used in incidence studies.

Perhaps the most common measure of disease occurrence is the person-time incidence rate (or hazard rate, force of mortality or incidence density (Miettinen, 1985)) which is a measure of the disease occurrence per unit population time, and has the reciprocal of time as its dimension. In this example (table 2.2), there were 952 cases of disease diagnosed in the non-exposed group during the ten years of follow-up, which involved a total of 95,163 person-years; this is less than the total possible person-time of 100,000 person-years since people who developed the disease before the end of the ten-year period were no longer "at risk" of developing it, and stopped contributing person-years at that time (for simplicity I have ignored the problem of people whose disease disappears and then reoccurs over time, and I have assumed that we are studying the incidence of the first occurrence of disease). Thus, the incidence rate in the non-exposed group (b/Y_0) was $952/95,163 =$

0.0100 (or 1000 per 100,000 person-years).

A second measure of disease occurrence is the incidence proportion or average risk which is the proportion of people who experience the outcome of interest at any time during the follow-up period (the incidence proportion is often called the cumulative incidence, but the latter term is also used to refer to cumulative hazards (Breslow and Day, 1987)). Since it is a proportion it is dimensionless, but it is necessary to specify the time period over which it is being measured. In this instance, there were 952 incident cases among the 10,000 people in the non-exposed group, and the incidence proportion (b/N_0) was therefore $952/10,000 = 0.0952$ over the ten year follow-up period. When the outcome of interest is rare over the follow-up period (e.g. an incidence proportion of less than 10%), then the incidence proportion is approximately equal to the incidence rate multiplied by the length of time that the population has been followed (in the example, this product is 0.1000 whereas the incidence proportion is 0.0952). I have assumed, for simplicity, that no-one or was lost to follow-up during the study period (and therefore stopped contributing person-years to the study). However, as noted above when this assumption is not valid (i.e. when a significant proportion of people have died or have been lost to follow-up), then the incidence proportion cannot be estimated directly, but must be estimated indirectly from the incidence rate (which takes into account that follow-up was not complete) or from life tables (which stratify on follow-up time).

A third possible measure of disease occurrence is the incidence odds (Greenland, 1987) which is the ratio of the number of people who experience the outcome (b) to the number of people who do not experience the outcome (d). As for the incidence proportion, the incidence odds is dimensionless, but it is necessary to specify the time period over which it is being measured. In this example, the incidence odds (b/d) is $952/9,048 = 0.1052$. When the outcome is rare over the follow-up period then the incidence odds is approximately equal to the incidence proportion. Once again, if loss to follow-up is significant, then the incidence odds cannot be estimated directly, but must be

estimated indirectly from the incidence rate (via the incidence proportion, or via life-table methods). The incidence odds is not very interesting or useful as a measure of disease occurrence, but it is presented here because the incidence odds is used to calculate the incidence odds ratio which is estimated in certain case-control studies (see below).

These three measures of disease occurrence all involve the same numerator: the number of incident cases of disease (b). They differ in whether their denominators represent person-years at risk (Y_0), persons at risk (N_0), or survivors (d).

Table 2.2

Findings from a hypothetical cohort study of 20,000 persons followed for 10 years

	Exposed	Non-exposed	Ratio
Cases	1,813 (a)	952 (b)	
Non-cases	8,187 (c)	9,048 (d)	
Initial population size	10,000 (N_1)	10,000 (N_0)	
Person-years	90,635 (Y_1)	95,163 (Y_0)	
Incidence rate	0.0200 (I_1)	0.0100 (I_0)	2.00
Incidence proportion (average risk)	0.1813 (R_1)	0.0952 (R_0)	1.90
Incidence odds	0.2214 (O_1)	0.1052 (O_0)	2.11

Measures of Effect in Incidence Studies

Corresponding to these three measures of disease occurrence, there are three principal ratio measures of effect which can be used in incidence studies. The measure of interest is often the rate ratio (incidence density ratio), the ratio of the incidence rate in the exposed group (a/Y_1) to that in the non-exposed group (b/Y_0). In the example in table 2.2, the incidence rates are 0.02 per person-year in the exposed group and 0.01 per person-year in the non-exposed group, and the rate ratio is therefore 2.00.

A second commonly used effect measure is the risk ratio (incidence proportion ratio or cumulative incidence ratio) which is the ratio of the incidence proportion in the exposed group (a/N_1) to that in the non-exposed group (b/N_0). In this example, the risk ratio is $0.1813/0.0952 = 1.90$. When the outcome is rare over the follow-up period the risk ratio is approximately equal to the rate ratio.

A third possible effect measure is the incidence odds ratio which is the ratio of the incidence odds in the exposed group (a/c) to that in the non-exposed group (b/d). In this example the odds ratio is $0.2214/0.1052 = 2.11$. When the outcome is rare over the study period the incidence odds ratio is approximately equal to the incidence rate ratio.

These three multiplicative effect measures are sometimes referred to under the generic term of relative risk. Each involves the ratio of a measure of disease occurrence in the exposed group

to that in the non-exposed group. The various measures of disease occurrence all involve the same numerators (incident cases), but differ in whether their denominators are based on person-years, persons, or survivors (people who do not develop the disease at any time during the follow-up period). They are all approximately equal when the disease is rare during the follow-up period (e.g. an incidence proportion of less than 10%). However, the odds ratio has been severely criticised as an effect measure (Greenland, 1987; Miettinen and Cook, 1981), and has little intrinsic meaning in incidence studies, but it is presented here because it is the standard effect measure in incidence case-control studies (see below).

Finally, it should be noted that an analogous approach can be used to calculate measures of effect based on differences rather than ratios, in particular the rate difference and the risk difference. Ratio measures are usually of greater interest in etiologic research, because they have more convenient statistical properties, and it is easier to assess the strength of effect and the possible role of various sources of bias when using ratio measures (Cornfield et al, 1951). Thus, I will concentrate on the use of ratio measures in the remainder of this text. However, other measures (e.g. risk difference, attributable fraction) may be of value in certain circumstances, such as evaluating the public health impact of a particular exposure, and I encourage readers to consult standard texts for a comprehensive review of these measures (e.g. Rothman and Greenland, 1998).

2.2. Incidence Case-Control Studies

Incidence studies are the most comprehensive approach to studying the causes of disease, since they use all of the information about the source population over the risk period. However, they are very expensive in terms of time and resources. For example, the hypothetical study presented in table 2.2 would involve enrolling 20,000 people and collecting exposure information (on both past and present exposure) for all of them. The same findings can be obtained more efficiently by using a case-control design.

An incidence case-control study involves studying all (or a sample) of the incident cases of the disease that occurred in the source population over the risk period, and a control group sampled from the

same population over the same period (the possible methods of sampling controls are described below).

Table 2.3 shows the data from a hypothetical case-control study, which involved studying all of the 2,765 incident cases which would have been identified in the full incidence study, and a sample of 2,765 controls (one for each case). Such a case-control study would achieve the same findings as the full incidence study, but would be much more efficient, since it would involve ascertaining the exposure histories of 5,530 people (2,765 cases and 2,765 controls) rather than 20,000. When the outcome under study is very rare, an even more remarkable gain in efficiency can be achieved with very little reduction in the precision of the effect estimate.

Table 2.3

Findings from a hypothetical incidence case-control study based on the cohort in table 2.2

	Exposed	Non-exposed	Odds Ratio
Cases	1,813 (a)	952 (b)	
Controls: from survivors (cumulative sampling)	1,313 (c)	1,452 (d)	2.11
from source population (case-cohort sampling)	1,383 (c)	1,383 (d)	1.90
from person-years (density sampling)	1,349 (c)	1,416 (d)	2.00

Measures of Effect in Incidence Case-Control Studies

In case-control studies, the relative risk is estimated using the odds ratio.

Suppose that a case-control study is conducted in the study population shown in table 2.2; such a study might involve all of the 2,765 incident cases and a group of 2,765 controls (table 2.3). The effect measure which the odds ratio obtained from this case-control study will estimate depends on the manner in which controls are selected. Once again, there are three main options (Miettinen, 1985; Pearce, 1993; Rothman and Greenland, 1998).

One option, called cumulative (or cumulative incidence) sampling, is to select controls from those who do not experience the outcome during the follow-up period, i.e. the survivors (those who did not develop the disease at any time during the follow-up period). In this instance, the ratio of exposed to non-exposed controls will estimate the exposure odds ($c/d = 8178/9048 = 1313/1452$) of the survivors, and the odds ratio obtained in the case-control study will therefore estimate the incidence odds ratio in the source population over the study period (2.11). Early presentations of the case-control approach usually assumed this context (Cornfield, 1951), and it was emphasised that the odds ratio was approximately equal to the risk ratio when the disease was rare.

It was later recognised that controls can be sampled from the entire source population (those at risk at the beginning of follow-up), rather than just from the survivors (those at risk at the end of follow-up). This approach which was previously used by Thomas (1972) and Kupper et al (1975), has more

recently been termed case-cohort sampling (Prentice, 1986), or case-base sampling (Miettinen, 1982). In this instance, the ratio of exposed to non-exposed controls will estimate the exposure odds in the source population of persons at risk at the start of follow-up ($N_1/N_0 = 10000/10000 = 1383/1383$), and the odds ratio obtained in the case-control study will therefore estimate the risk ratio in the source population over the study period (1.90). In this instance the method of calculation of the odds ratio is the same as for any other case-control study, but minor changes are needed in the standard methods for calculating confidence intervals and p-values to take into account that some cases may also be selected as controls (Greenland, 1986).

The third approach is to select controls longitudinally throughout the course of the study (Sheehe, 1962; Miettinen, 1976); this is sometimes described as risk-set sampling (Robins et al, 1986), sampling from the study base (the person-time experience) (Miettinen, 1985), or density sampling (Kleinbaum et al, 1982). In this instance, the ratio of exposed to non-exposed controls will estimate the exposure odds in the person-time ($Y_1/Y_0 = 90635/95613 = 1349/1416$), and the odds ratio obtained in the case-control study will therefore estimate the rate ratio in the study population over the study period (2.00).

Case-control studies have traditionally been presented in terms of cumulative sampling (e.g. Cornfield, 1951), but most case-control studies actually involve density sampling (Miettinen, 1976), often with matching on a time variable such as calendar time or age, and therefore estimate the rate ratio without the need for any rare disease assumption (Sheehe, 1962; Miettinen, 1976; Greenland and Thomas, 1982).

Example 2.2

Gustavsson et al (2001) studied the risk of myocardial infarction from occupational exposure to motor exhaust, other combustion products, organic solvents, lead, and dynamite. They identified first-time, nonfatal myocardial infarctions among men and women aged 45-70

years in Stockholm County from 1992-1994. They selected controls from the general population living in the same County during the same period (i.e. density matching), matched for sex, age, year, and hospital catchment area. The odds ratio (estimating the rate ratio) of myocardial

infarction was 2.11 (95% CI 1.23-3.60) among those highly exposed occupationally, and 1.42 (95% CI 1.05-1.92) in those moderately exposed, compared with persons not occupationally exposed to combustion products from organic material.

Summary

When a dichotomous outcome is under study (e.g. being alive or dead, or having or not having a disease) a fundamental distinction is between studies of incidence and studies of prevalence. Thus, four main types of studies can be identified: incidence studies, incidence case-control studies, prevalence studies, and prevalence case-control studies (Morgenstern and Thomas, 1993; Pearce, 1998). These various study types differ according to whether they involve incidence or prevalence data and whether or not they involve sampling on the basis of the outcome under study. Incidence studies involve collecting and analysing data on

the exposure and disease experience of the entire source population. They may resemble randomized trials, but they may involve additional problems of confounding because exposure has not been randomly assigned. The other potential study designs all involve sampling from the source population, and therefore may include additional biases arising from the sampling process (chapter 6). In particular, incidence case-control studies involve sampling on the basis of outcome, i.e. they usually involve all incident cases generated by the source population and a control group (of non-cases) sampled at random from the source population.

References

- Breslow NE, Day NE (1987). Statistical methods in cancer research. Vol II: The analysis of cohort studies. Lyon, France: IARC.
- Checkoway H, Pearce N, Kriebel D (2004). Research methods in occupational epidemiology. 2nd ed. New York: Oxford University Press.
- Cornfield J (1951). A method of estimating comparative rates from clinical data: applications to cancer of the lung, breast and cervix. JNCI 11: 1269-75.
- Greenland S (1986). Adjustment of risk ratios in case-base studies (hybrid epidemiologic designs). Stat Med 5: 579-84.
- Greenland S (1987). Interpretation and choice of effect measures in epidemiologic analyses. Am J Epidemiol 125: 761-8.
- Greenland S, Thomas DC (1982). On the need for the rare disease assumption in case-control studies. Am J Epidemiol 116: 547-53.
- Gustavsson P, Plato N, Hallqvist J, et al (2001). A population-based case-referent study of myocardial infarction and occupational exposure to motor exhaust, other combustion products, organic solvents, lead and dynamite. Epidemiol 12: 222-8.
- Kleinbaum DG, Kupper LL, Morgenstern H (1982). Epidemiologic research. Principles and quantitative methods. Belmont, CA: Lifetime Learning Publications.
- Kupper LL, McMichael AJ, Spirtas R (1975). A hybrid epidemiologic design useful in estimating relative risk. J Am Stat Assoc 70:524-8.
- Martinez FD, Wright AJ, Taussig LM, et al (1995). Asthma and wheezing in the first six years of life. New Engl J Med 332: 133-8.
- Miettinen OS (1976). Estimability and estimation in case-referent studies. Am J Epidemiol 103: 226-35.
- Miettinen OS, Cook EF (1981). Confounding: essence and detection. Am J Epidemiol 114: 593-603.
- Miettinen O (1982). Design options in epidemiologic research: an update. Scand J Work Environ Health 8(suppl 1): 7-14.
- Miettinen OS (1985). Theoretical epidemiology. New York: Wiley.
- Morgenstern H, Thomas D (1993). Principles of study design in environmental epidemiology. Environ Health Perspectives 101: S23-S38.
- Pearce N (1993). What does the odds ratio estimate in a case-control study? Int J Epidemiol 22: 1189-92.
- Pearce N (1998). The four basic epidemiologic study types. J Epidemiol Biostat 3: 171-7.
- Prentice RL (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. Biometrika 73: 1-11.
- Robins JM, Breslow NE, Greenland S (1986). Estimation of the Mantel-Haenszel variance consistent with both sparse-data and large-strata limiting models. Biometrics 42: 311-23.

Rothman KJ, Greenland S (1998).
Modern epidemiology. 2nd ed.
Philadelphia: Lippincott-Raven.

Sheehe PR (1962). Dynamic risk analysis
of matched pair studies of disease.
Biometrics 18: 323-41.

Thomas DB (1972). The relationship of
oral contraceptives to cervical
carcinogenesis. *Obstet Gynecol* 40:
508-18.

CHAPTER 3. Prevalence Studies

(In: Pearce N. A Short Introduction to Epidemiology. 2nd ed. Wellington, CPHR, 2005)

Incidence studies are ideal for studying events such as mortality or cancer incidence, since they involve collecting and analysing all of the relevant information on the source population and we can get better information on when exposure and disease occurred. However, incidence studies involve lengthy periods of follow-up and large resources, in terms of both time and funding, and it may be difficult to identify incident cases of non-fatal chronic conditions such as diabetes. Thus, in some settings (e.g. some developing countries) and/or for some

conditions (e.g. chronic non-fatal disease) prevalence studies are the only option. Furthermore, in some instances we may be more interested in factors which affect the current burden of disease in the population. Consequently, although incidence studies are usual preferable, there is also an important role for prevalence studies, both for practical reasons, and because such studies enable the assessment of the level of morbidity and the population "disease burden" for a non-fatal condition.

3.1. Prevalence Studies

The term prevalence denotes the number of cases of the disease under study existing in the source population at a particular time. This can be defined as point prevalence estimated at one point in time, or period prevalence which denotes the number of cases that existed during some time interval (e.g. one year).

The prevalence is a proportion, and the statistical methods for calculating a confidence interval for the prevalence are identical to those presented above for calculating a confidence interval for the incidence proportion (chapter 12).

In some instances, the aim of a prevalence study may simply be to compare the disease prevalence among

a specific population with that in other communities or countries. This may be done, for example, in order to discover differences in disease prevalence and to thus suggest possible risk factors for the disease. These further studies may involve testing specific hypotheses by comparing prevalence in subgroups of people who have or have not been exposed to a particular risk factor (e.g. as passive smoking) in the past.

Prevalence studies often represent a considerable saving in resources compared with incidence studies, since it is only necessary to evaluate disease prevalence at one point in time, rather than continually searching for incident cases over an extended period of time. On the other hand, this gain in efficiency

is achieved at the cost of greater risk of biased inferences, since it may be much more difficult to understand the temporal relationship between various exposures and the occurrence of disease. For example, an exposure that increases the risk of death in people with pre-existing

chronic heart disease will be negatively associated with the prevalence of heart disease (in people who are alive!), and will therefore appear to be 'protective' against heart disease in a prevalence study.

Example 3.1

The International Study of Asthma and Allergies in Childhood (ISAAC) (Asher et al, 1995; Pearce et al, 1993) involved a simple Phase I global asthma symptom prevalence survey and a more in-depth Phase II survey. The emphasis was on obtaining the maximum possible participation across the world in order to obtain a global overview of childhood asthma prevalence, and the Phase I questionnaire modules were designed to be simple and to require minimal resources to administer. In addition, a video questionnaire involving the audio-visual presentation of clinical signs and symptoms of asthma was developed in order to minimise translation problems. The population of interest was schoolchildren aged 6-7 years and 13-14

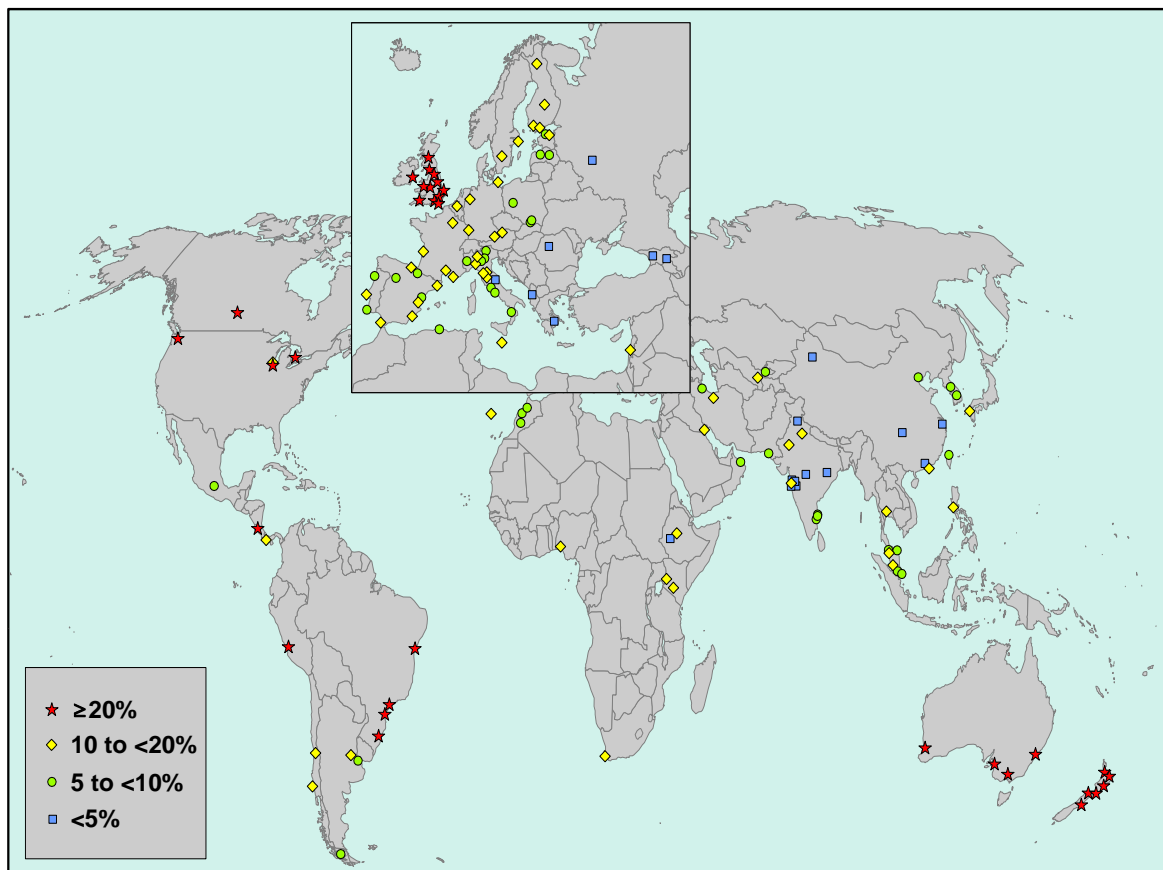
years within specified geographical areas. The older age-group was chosen to reflect the period when morbidity from asthma is common and to enable the use of self-completed questionnaires. The younger age-group was chosen to give a reflection of the early childhood years, and involves parent-completion of questionnaires. The Phase I findings, involving more than 700,000 children, showed striking international differences in asthma symptom prevalence (ISAAC Steering Committee, 1998a, 1998b). Figure 3.1 shows the findings for current wheeze (i.e. wheeze in the previous 12 months). There are a number of interesting features of the figure: (i) there is a particularly high prevalence of reported asthma

symptoms in English-speaking countries; (ii) centres in Latin America also had particularly high symptom prevalence; (iii) there is also high asthma prevalence in Western Europe, with lower prevalences in Eastern and Southern Europe - for example, there is a clear Northwest-Southeast gradient within Europe, with the highest prevalence in the world being in the United Kingdom, and some of the lowest prevalences in Albania and Greece; (iv) Africa and Asia generally showed relatively low asthma prevalence. These striking findings call into question many of the "established" theories of asthma causation, and have played a major role in the development of new theories of asthma causation in recent years (Douwes and Pearce, 2003).

Figure 3.1

Twelve month period prevalence of asthma symptoms in 13-14 year old children in Phase I of the International Study of Asthma and Allergies in Childhood (ISAAC)

Source: ISAAC Steering Committee (1998b)



Measures of Effect in Prevalence Studies

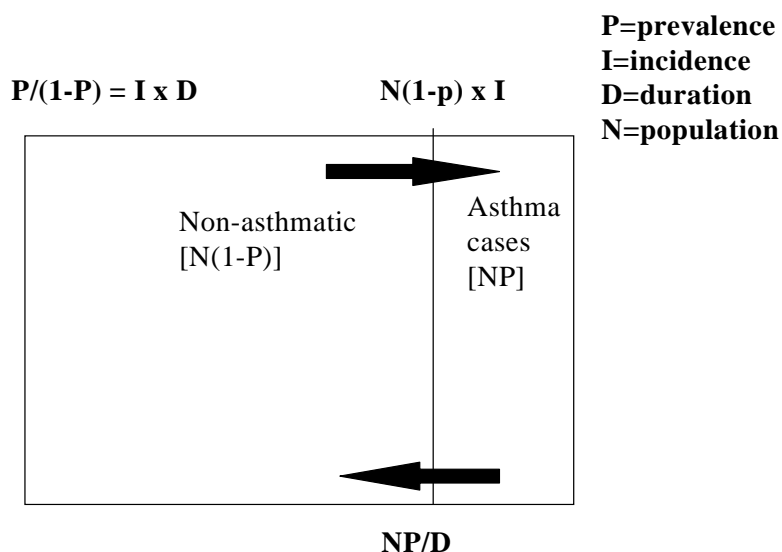
Figure 3.2 shows the relationship between incidence and prevalence of disease in a "steady state" population. Assume that the population is in a "steady state" (stationary) over time (in that the numbers within each subpopulation defined by exposure, disease and covariates do not change with time) – this usually requires that incidence rates and exposure and disease status are unrelated to the immigration and emigration rates and

population size - and that average disease duration (D) does not change over time. Then, if we denote the prevalence of disease in the study population by P, the prevalence odds is equal to the incidence rate (I) times the average disease duration (Alho, 1992):

$$\frac{P}{(1-P)} = ID$$

Figure 3.2

Relationship between prevalence and incidence in a "steady state" population



Now suppose that we compare two populations (indexed by 1=exposed and 0=non-exposed) and that both satisfy the above conditions. Then, the prevalence odds is directly proportional to the disease incidence, and the prevalence odds ratio (POR) satisfies the equation:

$$POR = \frac{[P_1/(1-P_1)]}{[P_0/(1-P_0)]} = \frac{I_1 D_1}{I_0 D_0}$$

An increased prevalence odds ratio may thus reflect the influence of factors that increase the duration of disease, as well as those that increase disease incidence. However, in the special case where the

average duration of disease is the same in the exposed and non-exposed groups (i.e. $D_1 = D_0$), then the prevalence odds ratio satisfies the equation:

$$POR = I_1/I_0$$

i.e. under the above assumptions, the prevalence odds ratio directly estimates the incidence rate ratio (Pearce, 2004). However, it should be emphasised that prevalence depends on both incidence and average disease duration, and a

difference in prevalence between two groups could entirely depend on differences in disease duration (e.g. because of factors which prolong or exacerbate symptoms) rather than differences in incidence. Changes in incidence rates, disease duration and population sizes over time can also bias the POR away from the rate ratio, as can migration into and out of the population at risk or the prevalence pool.

Table 3.1

Findings from a hypothetical prevalence study of 20,000 persons

	Exposed	Non-exposed	Ratio
Cases	909 (a)	476 (b)	
Non-cases	9,091 (c)	9,524 (d)	
Total population	10,000 (N_1)	10,000 (N_0)	
Prevalence	0.0909 (P_1)	0.0476 (P_0)	1.91
Prevalence odds	0.1000 (O_1)	0.0500 (O_0)	2.00

Table 3.1 shows data from a prevalence study of 20,000 people. This is based on the incidence study represented in table 2.2 (chapter 2), with the assumptions that, for both populations, the incidence rate and population size is constant over time, that the average duration of disease is five years, and that there is no migration of people with the disease into or out of the population (such assumptions may not be realistic, but are made here for purposes of illustration). In this situation, the number of cases who "lose" the disease each year is balanced by the

number of new cases generated from the source population. For example, in the non-exposed group, there are 476 prevalent cases, and 95 (20%) of these "lose" their disease each year; this is balanced by the 95 people who develop the disease each year (0.0100 of the susceptible population of 9524 people). With the additional assumption that the average duration of disease is the same in the exposed and non-exposed groups, then the prevalence odds ratio (2.00) validly estimates the incidence rate ratio (see table 2.2).

3.2. Prevalence Case-Control Studies

Just as an incidence case-control study can be used to obtain the same findings as a full incidence study, a prevalence case-control study can be used to obtain the same findings as a full prevalence study in a more efficient manner.

For example, in a prevalence study, obtaining exposure information may be difficult or costly, e.g. if it involves lengthy interviews, or expensive testing of biological samples. In this situation, a considerable gain in efficiency can be achieved by only obtaining exposure information on the prevalent cases and a sample of controls selected at random from the non-cases, rather than collecting exposure information for everyone in the prevalence study.

Measures of Effect in Prevalence Case-Control Studies

Suppose that a nested case-control study is conducted in the study population (table 3.1), involving all of the 1,385 prevalent cases and a group of 1,385 controls (table 3.2). The usual approach is to select controls from the non-cases. The ratio of exposed to non-exposed controls will then estimate the exposure odds (b/d) of the non-cases, and the odds ratio obtained in the prevalence case-control study will therefore estimate the prevalence odds ratio in the source population (2.00), which in turn estimates the incidence rate ratio provided that the assumptions described above are satisfied in the exposed and non-exposed populations.

Table 3.2

Findings from a hypothetical prevalence case-control study based on the population represented in table 3.1

	Exposed	Non-exposed	Ratio
Cases	909 (a)	476 (b)	
Controls	676 (c)	709 (d)	
Prevalence odds	1.34 (O_1)	0.67 (O_0)	2.00

Example 3.2

Studies of congenital malformations usually involve estimating the prevalence of malformations at birth (i.e. this is a prevalence rather than an incidence measure). Garcia et al (1999) conducted a (prevalence) case-control study of occupational

exposure to pesticides and congenital malformations in Comunidad Valenciana, Spain. A total of 261 cases and 261 controls were selected from those infants born in eight public hospitals during 1993-1994. For mothers who were involved in agricultural activities in

the month before conception and the first trimester of pregnancy, the adjusted prevalence odds ratio for congenital malformations was 3.2 (95% CI 1.1-9.0). There was no such association with paternal agricultural work.

Summary

When a dichotomous outcome is under study (e.g. being alive or dead, or having or not having a disease) four main types of studies can be identified: incidence studies, incidence case-control studies, prevalence studies, and prevalence case-control studies (Morgenstern and Thomas, 1993; Pearce, 1998). Prevalence studies involve measuring the prevalence of the disease in the source population at a

particular time, rather than the incidence of the disease over time. Prevalence case-control studies involve sampling on the basis of outcome, i.e. they usually involve all prevalent cases in the source population and a control group (of non-cases) sampled from the source population.

References

Alho JM (1992). On prevalence, incidence, and duration in general stable populations. *Biometrics* 48: 587-92.

Asher I, Keil U, Anderson HR, et al (1995). International study of asthma and allergies in childhood (ISAAC): rationale and methods. *Eur Resp J* 8: 483-91.

Douwes J, Pearce N (2003). Asthma and the Westernization "package". *Int J Epidemiol* 31: 1098-1102.

Garcia AM, Fletcher T, Benavides FG, Orts E (1999). Parental agricultural work and selected congenital malformations. *Am J Epidemiol* 149: 64-74.

- ISAAC Steering Committee (1998a).
Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis and atopic eczema: ISAAC. *Lancet* 351: 1225-32.
- ISAAC Steering Committee (1998b).
Worldwide variations in the prevalence of asthma symptoms: International Study of Asthma and Allergies in Childhood (ISAAC). *Eur Respir J* 12: 315-35.
- Morgenstern H, Thomas D (1993).
Principles of study design in environmental epidemiology. *Environ Health Perspectives* 101: S23-S38.
- Pearce N (1998). The four basic epidemiologic study types. *J Epidemiol Biostat* 3: 171-7.
- Pearce N (2004). Effect measures in prevalence studies. *Environmental Health Perspectives* 2004; 112: 1047-50.
- Pearce NE, Weiland S, Keil U, et al (1993). Self-reported prevalence of asthma symptoms in children in Australia, England, Germany and New Zealand: an international comparison using the ISAAC protocol. *Eur Resp J* 6: 1455-61.

CHAPTER 4. More Complex Study Designs

(In: Pearce N. A Short Introduction to Epidemiology. 2nd ed. Wellington, CPHR, 2005)

In the previous two chapters I reviewed the possible study designs for the simple situation where individuals are exposed to a particular risk factor (e.g. a particular chemical) and when a dichotomous outcome is under study (e.g. being alive or dead, or having or

not having a particular disease). I now consider studies involving other axes of classification, continuous measurements of health status (e.g. continuous lung function or blood pressure measurements) and more complex study designs (ecologic and multilevel studies).

4.1: Other Axes of Classification

The four basic study types discussed in chapters 2 and 3 are defined in terms of: (a) the type of outcome under study (incidence or prevalence); and (b) whether there is sampling on the basis of outcome. They do not involve any consideration of the nature of the exposure data. This provides additional axes of classification.

Continuous Exposure Data

Firstly, it should be noted that in discussing the above classification we have assumed that exposure is dichotomous (i.e. study participants are exposed or not exposed). In reality, there may be multiple exposure categories (e.g. high, medium and low exposure), or exposure may be measured as a continuous variable (see chapter 8). However, although this requires minor changes to the data analysis (see chapter 12), it does not alter the four-fold categorisation of study design options presented above.

The Timing of Collection Of Exposure Information

Perhaps the feature that has received the most attention in various classification schemes is the timing of the collection of exposure information. This has dominated discussions of "directionality", particularly with regard to case-control studies. In fact, for all of the four basic study types, exposure information can be collected prospectively or retrospectively. For example, an incidence study or incidence case-control study of occupational cancer may collect exposure information prospectively, or use historical information that was collected prospectively but abstracted retrospectively by the investigator (e.g. occupational hygiene monitoring records), or use exposure information that was collected retrospectively (e.g. recall of duration and intensity of pesticide use). An unfortunate aspect of some discussions of the merits of case-control studies is that they have often

been labelled as “retrospective” studies, when this is in fact not an inherent part of their design. The potential “problem” of bias due to exposure ascertainment errors (e.g. recall bias) arises from the retrospective collection of exposure information, irrespective of whether the study is an incidence, incidence case-control, prevalence, or prevalence case-control study.

Sources of Exposure Information

Another set of issues that occur in practice involve the sources of exposure information (e.g. routine records, job-exposure-matrices, questionnaires, biological samples). However, as noted above, these issues are important in understanding sources of bias but are

not fundamental to the classification of study types since, as with issues of directionality, they do not affect the parameterization of the exposure-outcome association.

The Level of Measurement of Exposure

A third additional axis of classification involves the level of measurement of exposure. In particular, in ecologic studies exposure information may be collected on a group rather than on individuals (e.g. average level of meat consumption) although others may still be available for individuals (e.g. age, gender). This situation is discussed in section 4.3.

4.2: Continuous Outcome Measures

Cross-Sectional Studies

In chapters 2 and 3, the health outcome under study was a state (e.g. having or not having hypertension). Studies could involve observing the incidence of the event of acquiring the disease state (e.g. the incidence of being diagnosed with hypertension), or the prevalence of the disease state (e.g. the prevalence of hypertension). More generally, the health state under study may have multiple categories (e.g. non-hypertensive, mild hypertension, moderate hypertension, severe hypertension) or may be represented by a continuous measurement (e.g. blood pressure). Since these measurements are taken at a particular point in time, such studies are often referred to as cross-sectional studies. Prevalence studies (see chapter 3) are a subgroup

of cross-sectional studies in which the disease outcome is dichotomous.

Although cross-sectional studies are sometimes described as studies in which exposure and disease information is collected at the same point in time (e.g. Kramer and Boivin, 1988; Last 1988), this is not in fact an inherent feature of such studies. In most cross-sectional studies (including prevalence studies), information on exposure will be physically collected by the investigator at the same time that information on disease is collected. Nonetheless, exposure information may include factors that do not change over time (e.g. gender) or change in a predictable manner (e.g. age) as well as factors that do change over time. The latter may have been measured at the time of data collection (e.g. current levels of airborne

dust exposure), or at a previous time (e.g. from historical records on past exposure levels) or integrated over time. The key feature of cross-sectional studies is that they involve studying disease at a particular point in time. Exposure information can be collected for current and/or historical exposures, and a wide variety of exposure assessment methods can be used within this general category of study (these are discussed further in chapter 8).

Just as a prevalence case-control study can be based on a prevalence survey, a cross-sectional study can also involve sampling on the basis of the disease outcome. For example, a cross-sectional study of bronchial hyperresponsiveness (BHR) could involve testing all study participants for BHR and then categorising the test results into severe BHR, mild BHR, and no BHR, and then obtaining exposure information on all severe BHR cases and from random samples of the other two groups.

Measures of Effect in Cross-Sectional Studies

In a simple cross-sectional study involving continuous outcome data, the basic methods of statistical analysis involve comparing the mean level of the outcome in "exposed" and "non-exposed" groups, e.g. the mean levels of blood pressure in "exposed" and "non-exposed" people. Standard statistical methods of analysis for comparing means (perhaps after a suitable transformation to normalise the data), and calculating confidence intervals (and associated p-values) for differences between means, can be used to analyse such studies (see chapter 12). More generally, regression methods can be used to model the relationship between the level of exposure (measured as a continuous variable) and the level of the outcome measure (also measured as a continuous variable) (e.g. Armitage et al, 2002).

Example 4.1

Nersesyan et al (2001) studied chromosome aberrations in lymphocytes of persons exposed to an earthquake in Armenia. They collected blood samples from 41 victims of the 1988 earthquake and from 47 reference blood donors. Those

"exposed" to the earthquake had a higher proportion of cells with chromosome aberrations (3.1% (SD 2.1)) than the referents (1.7% (SD 1.3)). The differences persisted when the data were adjusted for age and gender. The authors suggested that the

findings could be due either to environmental exposures related to the earthquake or to severe psychogenic stress. They noted that studies in wild rodents living in seismic regions have shown similar findings.

Longitudinal Studies

Longitudinal studies (cohort studies) involve repeated observation of study participants over time (Pearce et al, 1998). Incidence studies (chapter 2) are a subgroup of longitudinal study in which the outcome measure is dichotomous. More generally, longitudinal studies may involve repeated assessment of categorical or continuous outcome measures over time (e.g. a series of linked cross-sectional studies in the same population). They thus can involve incidence data, a series of prevalence surveys, or a series of cross-sectional continuous outcome measures.

General longitudinal studies

A simple longitudinal study may involve comparing the disease outcome measure, or more usually changes in the measure over time, between exposed and non-exposed groups. For example, rather than comparing the incidence of hypertension (as in an incidence study),

or the prevalence at a particular time (as in a prevalence study), or the mean blood pressure at a particular point in time (as in a cross-sectional study), a longitudinal study might involve measuring baseline blood pressure in exposed and non-exposed persons and then comparing changes in mean blood pressure (i.e. the change from the baseline measure) over time in the two groups. Such a comparison of means can be made using standard statistical methods for comparing means and calculating confidence intervals and associated p-values for the difference between the means (Armitage et al, 2002; Beaglehole et al, 1993). More generally, regression methods (Diggle et al, 1994) might be used to model the relationship between the level of exposure (measured as a continuous variable) and the level of the outcome measure (also measured as a continuous variable, in this instance the change in FEV₁).

Example 4.2

The Tokelau Island Migrant Study (Wessen et al, 1992) examined the effects of migration on development of 'Western diseases' within a population which initially had a low incidence of these conditions. Round I surveys were conducted in the Tokelau Islands in 1968/1971, and these

were repeated (Round II) in both the Tokelau Islands (1976) and in New Zealand (1975-7). A regression analysis of changes in blood pressure between Round I and Round II (adjusted for age) found that the mean annual increase in blood pressure was greater in those who had migrated than in those

who had not: the mean differences were 1.43 for systolic and 1.15 for diastolic in men, and 0.66 and 0.46 respectively in women. These differences in rates of annual increase in blood pressure were maintained in subsequent surveys in men, but not in women.

Time series

One special type of longitudinal study is that of "time series" comparisons in which variations in exposure levels and symptom levels are assessed over time with each individual serving as their own control. Thus, the comparison of "exposed" and "non-exposed" involves the same persons evaluated at different times, rather than different groups of persons being compared (often at the same time) as in other longitudinal studies. The advantage of the time series approach is that it reduces or eliminates confounding (see chapter 6) by factors which vary among subjects but not over time (e.g. genetic factors), or whose day to day variation is unrelated to the main exposure (Pope and Schwartz, 1996). On the other hand, time series data often require special statistical techniques because any two factors that show a time trend will be correlated (Diggle et al, 1994). For example, even a three-month study of lung function in children will generally show an upward trend due to growth, as well as learning effects (Pope and Schwartz, 1996). A further problem is that the change in a measure over time may depend on the baseline value, e.g. changes in lung function over time may depend on the baseline level (Schouten and Tager, 1996).

Time series can involve dichotomous (binary) data, continuous data, or "counts" of events (e.g. hospital admissions) (Pope and Schwartz, 1996), and the changes in these values may be

measured over minutes, hours, days, weeks, months or years (Dockery and Brunekreef, 1996). In many instances, such data can be analysed using the standard statistical techniques outlined above. For example, a study of daily levels of air pollution and asthma hospital admission rates can be conceptualised as a study of the incidence of hospital admission in a population exposed to air pollution compared with that in a population not exposed to air pollution. The key difference is that only a single population is involved, and it is regarded as exposed on high pollution days and as non-exposed on low pollution days. Provided that the person-time of exposure is appropriately defined and assessed, then the basic methods of analysis are not markedly different from other studies involving comparisons of exposed and non-exposed groups.

However, the analysis of time series may be complicated because the data for an individual are not independent and serial data are often correlated (Sherrill and Viegli, 1996), i.e. the value of a continuous outcome measure on a particular day may be correlated with the value for the previous day. Furthermore, previous exposure may be as relevant as, or more relevant than, current exposure. For example, the effects of air pollution may depend on exposure on preceding days as well as on the current day (Pope and Schwartz, 1996).

Example 4.3

Hoek et al (2001) studied associations between daily variations in air pollution and mortality in The Netherlands during 1986-1994. The authors found (table 4.1) that heart disease deaths were increased during periods with high levels of ozone, black smoke, particulate matter 10 microns in diameter (PM₁₀), carbon monoxide (CO), sulfur dioxide

(SO₂) and nitrogen dioxide (NO₂). As with previously published studies, the effects depended on exposures on the previous few days, and were weaker when the analysis only considered exposures on a particular day without using any lag period (Schwartz, 2000). The authors reported that deaths due to heart failure, arrhythmia, cerebrovascular causes

and thrombotic causes were more strongly associated with air pollution than were cardiovascular deaths in general. In particular, heart failure deaths, which made up 10% of all cardiovascular deaths, were responsible for about 30% of the excess cardiovascular deaths related to air pollution from particulate matter, SO₂, CO, and NO₂.

Table 4.1

Relative risks* (and 95% CIs) of cardiovascular disease mortality associated with air pollution concentrations in the Netherlands

Pollutant	Total CVD mortality	Heart failure mortality
Ozone (1 day lag)	1.055 (1.032-1.079)	1.079 (1.009-1.154)
Black smoke (7 day mean)	1.029 (1.013-1.046)	1.081 (1.031-1.134)
PM ₁₀ (7 day mean)	1.012 (0.984-1.041)	1.036 (0.960-1.118)
CO (7 day mean)	1.026 (0.993-1.060)	1.109 (1.012-1.216)
SO ₂ (7 day mean)	1.029 (1.012-1.046)	1.098 (1.043-1.156)
NO ₂ (7 day mean)	1.023 (1.009-1.036)	1.064 (1.024-1.106)

*Relative risks per 1 to 99th percentile pollution difference
 Relative risks per 150 $\mu\text{g}/\text{m}^3$ for ozone (8-hour maximum of the previous Day), per 120 $\mu\text{g}/\text{m}^3$ for CO, per 80 $\mu\text{g}/\text{m}^3$ for PM₁₀, per 30 $\mu\text{g}/\text{m}^3$ for NO₂, and per 40 $\mu\text{g}/\text{m}^3$ for black smoke and SO₂, all as 7-day moving averages

Source: Hoek et al (2001)

4.3 Ecologic and Multilevel Studies

The basic study designs described in chapters 2 and 3 involved the measurement of exposure and disease in individuals. In this section, I consider more complex study designs in which exposures are measured in populations instead of, or in addition to, individuals.

Ecologic Studies

In ecologic studies exposure information may be collected on a group rather than on individuals. In the past, ecologic studies have been regarded as an inexpensive but unreliable method for studying individual-level risk factors for disease. For example, rather than go to the time and expense to establish a cohort study or case-control study of fat intake and breast cancer, one could simply use national dietary and cancer incidence data and, with minimal time and expense, show a strong correlation internationally between fat intake and breast cancer. In this situation, an ecologic study does not represent a fundamentally different study design, but merely a particular variant of the four basic study designs described in chapter 2 in which information on average levels of exposure in populations is used as a surrogate measure of exposure in individuals.

This approach has been quite rightly regarded as inadequate and unreliable because of the many additional forms of bias that can occur in such studies compared with studies of individuals within a population. In particular, not only will measures of exposure in populations often be poor surrogates for exposures in individuals, but the

'ecologic fallacy' (see below) can occur in that factors that are associated with national disease rates may not be associated with disease in individuals (Greenland and Robins, 1994). Thus, ecologic studies have recently been regarded as a relic of the "pre-modern" phase of epidemiology before it became firmly established with a methodologic paradigm based on the theory of randomized controlled trials of individuals.

However, population-level studies are now experiencing a revival for two important reasons (Pearce, 2000).

Firstly, it is increasingly recognised that, even when studying individual-level risk factors, population-level studies play an essential role in defining the most important public health problems to be addressed, and in generating hypotheses as to their potential causes. Many important individual-level risk factors for disease simply do not vary enough within populations to enable their effects to be identified or studied (Rose, 1992). More importantly, such studies are a key component of the continual cycle of theory and hypothesis generation and testing (Pearce, 2000).

Historically, the key area in which epidemiologists have been able to "add value" has been through this population focus (Pearce, 1996, 1999). For example, many of the recent discoveries on the causes of cancer (including dietary factors and colon cancer, hepatitis B and liver cancer, aflatoxins and liver cancer, human papilloma virus and cervical cancer) have their origins, directly or indirectly, in the systematic international comparisons of cancer

incidence conducted in the 1950s and 1960s (Doll et al, 1966). These suggested hypotheses concerning the possible causes of the international patterns, which were investigated in more depth in further studies. In some instances these hypotheses were

consistent with biological knowledge at the time, but in other instances they were new and striking, and might not have been proposed, or investigated further, if the population level analyses had not been done.

Example 4.4

The International Study of Asthma and Allergies in Childhood (ISAAC) (Asher et al, 1995; Pearce et al, 1993) was described in example 3.1. Figure 4.1 shows the findings for current wheeze (i.e. wheeze in the previous 12 months) and its association with

tuberculosis notification rates (von Mutius et al, 2000). It shows a negative association between tuberculosis rates and asthma prevalence. This is not compelling evidence in itself (because of the major shortcoming of ecologic analyses that

are described below), but it is generally consistent with the "hygiene hypothesis" that suggests that asthma prevalence is increasing in Western countries because of the loss of a protective effect from infections such as tuberculosis in early life.

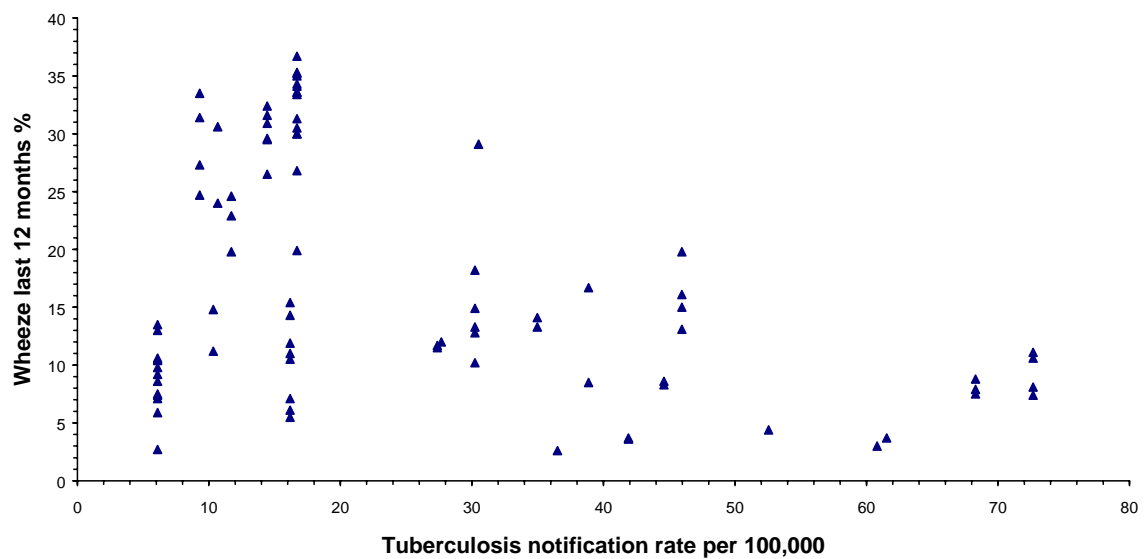
A second reason that ecologic studies are experiencing a revival is that it is increasingly being recognised that some risk factors for disease genuinely operate at the population level (Pearce, 2000). In some instances they may directly cause disease, but perhaps more commonly they may cause disease as effect modifiers or determinants of exposure to individual-level risk factors. For example, being poor in a rich country or neighbourhood may be worse than having the same income level in a poor country or neighbourhood, because of problems of social exclusion and lack of access to services and resources (Yen

et al, 1999). The failure to take account of the importance of population context, as an effect modifier and determinant of individual-level exposures could be termed the "individualistic fallacy" (Diez-Roux, 1998) in which the major population determinants of health are ignored and undue attention is focussed on individual characteristics. In this situation, the associations between these individual characteristics and health can be validly estimated, but their importance relative to other potential interventions, and the importance of the context of such interventions, may be ignored.

Figure 4.1

Association of tuberculosis notification rates for the period 1980-1982 (in countries with valid tuberculosis notification data) and the prevalence of asthma symptoms in 13-14 year old children in the International Study of Asthma and Allergies in Childhood (ISAAC)

Source: von Mutius et al (2000)



Example 4.5

Wilkinson (1992) has analysed measures of income inequality and found them to be positively associated with national mortality rates in a number of Western countries. This is a true "ecologic exposure" since the level of income inequality is a characteristic of a country, and not of an individual. If this evidence is correct, this

is clearly of crucial importance since it implies that 'development' in itself may not automatically be good for health, and that the way in which the Gross National Product (GNP) is 'shared' may be as important as its absolute level. It should be noted, however, that this evidence has been disputed by other

researchers (e.g. Lynch et al, 2000; Pearce and Davey Smith, 2003) who have argued that the level of income inequality in a country, or in a state, is a surrogate measure for other socioeconomic factors, including the provision of public education and health services, as well as social welfare services.

Ecologic Fallacies

While stressing the potential value of ecologic analyses, it is also important to recognise their limitations. In particular, ecologic studies are a very poor means of assessing the effects of individual exposures (e.g. diet or tobacco smoking) since confounding (and effect modification) can occur at the individual level, the country (population) level, or both (Morgenstern, 1998). For example, almost any disease that is associated with affluence and westernisation has in the past been associated at the national level with sales of television sets, and nowadays is probably associated at the national level with rates of internet

usage. This does not mean that watching television causes every type of disease, but rather than in many instances the association between sales of television sets and disease at the national level is confounded by other exposures (at both the national and individual level). A hypothetical example is given in example 4.6. Another problem is that individual level effects can confound ecologic estimates of population-level effects (Greenland, 2001). These problems of cross-level inference are avoided (or reduced) in multilevel analyses (see below).

Example 4.6

Table 4.2 shows the data for a hypothetical ecological analysis. The numbers of cases and population numbers (and hence disease rates), as well as the percentage of the population exposed, are known for each country. Thus, the numbers of people exposed and non-exposed within each country are known, but it is not known how many cases were exposed and how many were not; thus it is not possible to estimate the rates in the exposed and non-exposed groups within each country. The country-level data indicate a negative association between exposure and disease at

the country level: if a regression is performed on the country-level data it indicates (comparing 100% exposure with 0% exposure) a relative risk of 0.5. However, it is not known whether this association applies to individuals, since the data are not available.

Tables 4.3-4.5 give three different scenarios, each of which could generate the data in table 4.2. In table 4.3, there is no confounding at the country level (because the rate in the non-exposed is the same - 200 per 1,000 - in each country), although there could of course still be uncontrolled confounding at the individual level.

Thus, the ecologic analysis correctly estimates the individual-level relative risk of 0.5. In table 4.4, there is confounding at the country level (because the rate in the non-exposed differs by country) and there is in fact no association at the individual level. In table 4.5, there is effect modification at the country level, and the relative risk is positive, but of differing magnitude, in all three countries. These three very different situations (a protective effect, no effect, a positive effect which is different in each country) all yield the same country-level data shown in table 4.2.

Table 4.2

Hypothetical example of an ecologic analysis

	Country 1 (35% exposed)		Country 2 (50% exposed)		Country 3 (65% exposed)	
	Cases	Rate	Cases	Rate	Cases	Rate
Exposed	?/ 7000	?	?/ 10000	?	?/ 13000	?
Non-exposed	?/ 13000	?	?/ 10000	?	?/ 7000	?
Total	33/ 20000	165	30/ 20000	150	27/ 20000	135

Source: Adapted from Morgenstern (1998)

Table 4.3

Hypothetical example of an ecologic analysis:
No confounding by country

	Country 1 (35% exposed)		Country 2 (50% exposed)		Country 3 (65% exposed)	
	Cases	Rate	Cases	Rate	Cases	Rate
Exposed	7/ 7000	100	10/ 10000	100	13/ 13000	100
Non-exposed	26/ 13000	200	20/ 10000	200	14/ 7000	200
Total	33/ 20000	165	30/ 20000	150	27/ 20000	135
Ratio		0.5		0.5		0.5

Source: Adapted from Morgenstern (1998)

Table 4.4

Hypothetical example of an ecologic analysis:
Confounding by country

	Country 1 (35% exposed)		Country 2 (50% exposed)		Country 3 (65% exposed)	
	Cases	Rate	Cases	Rate	Cases	Rate
Exposed	12/ 7000	171	15/ 10000	150	18/ 13000	139
Non-exposed	21/ 13000	162	15/ 10000	150	9/ 7000	129
Total	33/ 20000	165	30/ 20000	150	27/ 20000	135
Ratio	1.1		1.0		1.1	

Source: Adapted from Morgenstern (1998)

Table 4.5

Hypothetical example of an ecologic analysis:
Effect modification by country

	Country 1 (35% exposed)		Country 2 (50% exposed)		Country 3 (65% exposed)	
	Cases	Rate	Cases	Rate	Cases	Rate
Exposed	20/ 7000	286	20/ 10000	200	20/ 13000	154
Non-exposed	13/ 13000	100	10/ 10000	100	7/ 7000	100
Total	33/ 20000	165	30/ 20000	150	27/ 20000	135
Ratio	2.9		2.0		1.5	

Source: Adapted from Morgenstern (1998)

Multilevel Studies

If individual as well as population-level data are available, then the problems of cross-level confounding and effect modification (illustrated in example 4.6) are avoided by using multilevel modelling (Greenland, 2000, 2002). This enables the simultaneous consideration of individual level effects (e.g. individual income) and population-level effects (e.g. per capita national income, or income inequality). This approach therefore combines the

best features of individual level analyses and population-level analyses. In particular, it enables us to take the population context of exposure into account (Pearce, 2000). However, it should be stressed that multilevel modelling is complex, and requires intensive consideration of possible biases at the population level, as well as at the individual level (Blakely and Woodward, 2000).

Example 4.7

Yen and Kaplan (1999) conducted a multi-level analysis of neighbourhood social environment and risk of death in the Alameda County Study, comprising 6,928 non-institutionalised adult residents of the County

recruited in 1965. Mortality risks were significantly higher in neighbourhoods with a "low social environment", even after account was taken of individual income level, education, ethnicity,

perceived health status, smoking status, body mass index, and alcohol consumption. The authors concluded that the findings demonstrate the importance of area characteristics as a health risk factor.

Summary

The basic study designs presented in chapters 2 and 3 can be extended in two ways: by the inclusion of continuous outcome measures; and by the use of exposure information on populations rather than individuals.

Cross-sectional studies can include a variety of measurements of the health outcome under study (e.g. lung function or blood pressure measurements).

Prevalence studies are a subgroup of cross-sectional studies in which the outcome measure is dichotomous. Similarly, longitudinal studies can involve incidence data, but may also involve a series of cross-sectional measurements. Incidence studies are a subgroup of longitudinal studies in which the outcome measure is dichotomous. Time series studies are a particular type of longitudinal study in

which each subject serves as his or her own control.

Ecologic studies play an important role in the process of hypothesis generation and testing, but they pose additional problems of bias when attempting to

estimate the effects of exposures in individuals. These problems are avoided (or reduced) in multilevel analyses, which permit us to take the population context of exposure into account.

References

- Armitage P, Berry G, Matthews JNS (2002). *Statistical methods in medical research*. 4th ed Oxford: Blackwell.
- Asher I, Keil U, Anderson HR, Beasley R, et al (1995). International study of asthma and allergies in childhood (ISAAC): rationale and methods. *Eur Resp J* 8: 483-91.
- Beaglehole R, Bonita R, Kjellstrom T (1993). *Basic epidemiology*. Geneva: WHO.
- Blakeley T, Woodward AJ (2000). Ecological effects in multi-level studies. *J Epidemiol Comm Health* 54: 367-74.
- Diez-Roux AV (1998). Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *Am J Publ Health* 1998; 88: 216-22.
- Diggle PJ, Liang K-Y, Zeger SL (1994). *Analysis of longitudinal data*. Oxford: Clarendon.
- Dockery DW, Brunekreef B (1996). Longitudinal studies of air pollution effects on lung function. *Am J Respir Crit Care Med* 154: S250-S256.
- Doll R, Payne P, Waterhouse J (eds) (1966). *Cancer Incidence in Five Continents: A Technical Report*. Berlin: Springer-Verlag (for UICC).
- Greenland S, Robins J (1994). Ecologic studies - biases, misconceptions, and counterexamples. *Am J Epidemiol* 139: 747-60.
- Greenland S (2000). Principles of multilevel modelling. *Int J Epidemiol* 2000; 29: 158-67.
- Greenland S (2001). Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects, *Int J Epidemiol* 30: 1343-50.
- Greenland S (2002). A review of multilevel theory for ecologic analyses. *Stat Med* 21: 389-95.
- Hoek G, Brunekreef B, Fischer P, van Mijnen J (2001). The association between air pollution and heart failure, arrhythmia, embolism, thrombosis, and other cardiovascular causes of death in a time series study. *Epidemiol* 2001; 12: 355-57.
- Kramer MS, Boivin J-F (1988). The importance of directionality in epidemiologic research design. *J Clin Epidemiol* 41: 717-8.
- Last JM (ed) (1988). *A dictionary of epidemiology*. New York: Oxford University Press.
- Lynch J, Due P, Muntaner C, Davey Smith G (2000). Social capital – is it a good investment strategy for public

- health? *J Epidemiol Comm Health* 54: 404-8.
- Morgenstern H (1998). Ecologic studies. In: Rothman K, Greenland S. *Modern epidemiology*. Philadelphia: Lippincott-Raven, pp 459-80.
- Nersesyan AK, Boffetta P, Sarkisyan TF, et al (2001). Chromosome aberrations in lymphocytes of persons exposed to an earthquake in Armenia. *Scand J Work Environ Health* 27: 120-4.
- Pearce N (1996). Traditional epidemiology, modern epidemiology, and public health. *AJPH* 1996; 86: 678-83.
- Pearce N (1999). Epidemiology as a population science. *Int J Epidemiol* 1999; 28: S1015-8.
- Pearce N (2000). The ecologic fallacy strikes back. *J Epidemiol Comm Health* 2000; 54: 326-7.
- Pearce N, Davey Smith G (2003). Is social capital the key to inequalities in health? *Am J Publ Health* 93: 122-9.
- Pearce NE, Weiland S, Keil U, et al (1993). Self-reported prevalence of asthma symptoms in children in Australia, England, Germany and New Zealand: an international comparison using the ISAAC protocol. *Eur Resp J* 6: 1455-61.
- Pearce N, Beasley R, Burgess C, Crane J (1998). *Asthma epidemiology: principles and methods*. New York: Oxford University Press.
- Pearce N, Douwes J, Beasley R (2000). The rise and rise of asthma: a new paradigm for the new millenium? *J Epidemiol Biostat* 2000; 5: 5-16.
- Pope CA, Schwartz J (1996). Time series for the analysis of pulmonary health data. *Am J Respir Crit Care Med* 154: S229-S233.
- Rose G (1992). *The strategy of preventive medicine*. Oxford: Oxford University Press, 1992.
- Schouten JP, Tager IB (1996). Interpretation of longitudinal studies: an overview. *Am J Respir Crit Care Med* 154: S278-S284.
- Schwartz J (2000). The distributed lag between air pollution and daily deaths. *Epidemiol* 2000; 11: 320-6.
- Sherill D, Viegi G (1996). On modeling longitudinal pulmonary function data. *Am J Respir Crit Care Med* 154: S217-S222.
- Von Mutius E, Pearce N, Beasley R, Cheng S, Von Ehrenstein O, Björkstén B, Weiland S, on behalf of the ISAAC Steering Committee (2000). International patterns of tuberculosis and the prevalence of symptoms of asthma, rhinitis and eczema. *Thorax* 55: 449-53.
- Wessen AF, Hooper A, Huntsman J, et al (1992). Migration and health in a small society: The case of Tokelau. Oxford: Clarendon Press, 1992, pp 318-57.
- Wilkinson RG (1992). Income distribution and life expectancy. *Br Med J* 304: 165-8.
- Yen IH, Kaplan GA (1999). Neighbourhood social environment and risk of death: multilevel evidence from the Alameda County Study. *Am J Epidemiol* 149: 898-907.

Part II

Study Design Issues

CHAPTER 5: Precision

(In: Pearce N. A Short Introduction to Epidemiology. 2nd ed. Wellington, CPHR, 2005)

Random error will occur in any epidemiologic study, just as it occurs in experimental studies. It is often referred to as chance, although it can perhaps more reasonably be regarded as "ignorance" (although it is not the only thing that we may be ignorant about as our study may be biased by unknown confounders, measurement error, etc). For example, if we toss a coin 50 times, then ideally we might be able to predict the outcome of each "toss" based on the speed, spin, and trajectory of the coin. In practice, we do not have all of the necessary information (because of "ignorance"), or the computing power to use it (because of chaotic behaviour), and we therefore regard the outcome of each "toss" as a "chance" phenomenon. However, we may note that, on the average, 50% of the "tosses" are heads and therefore we may say that a particular toss has a "50% chance" of producing a head.

Similarly, suppose that 50 lung cancer deaths occurred among 10,000 people aged 35-39 exposed to a particular factor during one year. Then, if each person had exactly the same cumulative exposure, we might expect two subgroups of 5,000 people each to experience 25 deaths during the one-year period. However, just as 50 tosses of a coin will not usually produce exactly 25 heads and 25 tails, neither will there be exactly 25 deaths in each group. This occurs because of differences in exposure to other risk factors for lung cancer, and differences in individual susceptibility between the two groups. Ideally, we should attempt to gather information on all known risk factors (potential confounders), and to adjust for these in the analysis (see chapter

12). However, there will always be other unknown or unmeasurable risk factors operating, and hence the disease rates in particular subgroups will fluctuate about the average. This will occur even if each subgroup has exactly the same exposure history.

Even in an experimental study, in which participants are randomised into "exposed" and "non-exposed" groups, there will be "random" differences in background risk between the compared groups, but these will diminish in importance (i.e. the random differences will tend to "even out") as the study size grows. In epidemiological studies, because of the lack of randomisation, there is no guarantee that differences in baseline (background) risk will "even out" between the exposure groups as the study size grows.

The basic principles of analysis of epidemiologic data are discussed in chapter 12. However, at this stage it is important to discuss some basic statistical principles and methods since they are relevant to the calculation of the appropriate study size.

5.1: Basic Statistics

Basic Concepts

Data can be summarized in various forms, including frequency tables, histograms, bar charts, cross-tabulations and pie charts. However, it is usually also useful to give a summary measure of central tendency. The mean (or average) is the most commonly used measure of central tendency, because of its convenient statistical properties. The next step is data smoothing which involves the combination of the data with a statistical model. In the simplest case, this involves assuming a particular statistical distribution in order to obtain a summary measure of variability of the data. The most common measure of variability is the standard deviation (Armitage et al, 2002). The standard deviation is especially useful when the underlying data distribution is approximately normal (i.e. symmetric with a special type of bell-shape). If data is not normally distributed, then it can often be made approximately normally distributed by an appropriate transformation (e.g. a log transformation), but these transformations may distort the scientific meaning of the findings, and make them difficult to interpret.

Usually it is not possible to study the entire population in which one is interested (theoretically, this is almost always infinite since we usually wish to generalise our findings not only to the population we are studying, but also to other populations). It is therefore necessary to consider a random sample and to relate its characteristics to the total population. If repeated samples are

taken from the same population, then the mean will vary between samples. Even if the underlying population is not normally distributed, the means of the samples will be approximately normally distributed provided that the samples are sufficiently large (how "large" depends on how non-normally distributed the population is). The standard deviation of the sample means is termed the standard error of the mean. Since the means are approximately normally distributed, about 95% of sample means will lie within 1.96 standard errors of the overall population mean. Usually, a study only involves one sample, but the standard error can be estimated by dividing the standard deviation of the sample by the square root of the number of people in the sample.

Most epidemiological studies involve categorical rather than continuous outcome data. For example, in a particular area one might estimate the proportion of births involving congenital malformations over a particular time period (this is actually the prevalence at birth - it is very difficult to calculate the incidence of congenital malformations because this requires information on abortions and stillbirths as well as live births). This involves the calculation of a proportion (p). Under the binomial distribution, if the sample is sufficiently large, the sampling distribution will approximate to the normal distribution with mean (p) and standard deviation:

$$s = (p(1-p)/n)^{0.5}$$

where the "^{0.5}" indicates the square root of the expression in parentheses. Thus,

one can calculate the proportion with malformations (i.e. the mean score for a population in which a malformation scores 1 and a completely healthy baby scores 0), and the standard deviation of this proportion (i.e. the standard error of the mean score), and if the sample is sufficiently large one can analyze these estimates based on the normal distribution.

Testing and Estimation

Usually, in epidemiologic studies, we wish to measure the difference in disease occurrence between groups exposed and not exposed to a particular factor. For example, if we have estimated the proportion of pregnancies involving congenital malformations in an area with high nitrate levels in drinking water, then we would wish to compare this to the corresponding proportion in an area with low nitrate levels (or with

the proportion in all births nationally. In doing so, we not only wish to estimate the size of the observed association, but also whether an association as large as this is likely to have arisen by chance, if in fact there is no causal association between exposure and disease. The p-value is the probability that differences as large or larger as those observed could have arisen by chance if the null hypothesis (of no association between exposure and disease) is correct. In the past, it is been common to "test" the statistical significance of the study findings by seeing whether the p-value is less than an arbitrary value (e.g. $p < 0.05$). The limitations of statistical significance testing are discussed in chapter 12. However, even if we do not intend to use p-values when reporting the findings of a study, the statistical principles involved are nevertheless relevant to determining the appropriate study size.

5.2: Study Size and Power

The most effective means of reducing random error is by increasing the study size, so that the precision of the measure of association (the effect estimate) will be increased, i.e. the confidence intervals will be narrower. Random error thus differs from systematic error (see chapter 6) which cannot be reduced simply by increasing the study size.

A second factor that can affect precision, given a fixed total study size, is the relative size of the reference group (the unexposed group in a cohort study, or the controls in a case-control study). When exposure is not associated with disease (i.e. the true relative risk is 1.0), and the costs (of recruitment, data

collection, etc) of index and reference subjects are the same, then a 1:1 ratio is most efficient for a given total study size. When exposure increases the risk of the outcome, or referents are cheaper to include in the study than index subjects, then a larger ratio may be more efficient. The optimal reference: index ratio is rarely greater than 2:1 for a simple unstratified analysis (Walter, 1977) with equal index and referent costs, but a larger average ratio may be desirable in order to assure an adequate ratio in each stratum for stratified analyses.

The ideal study would be infinitely large, but practical considerations set limits on

the number of participants that can be included. Given these limits, it is desirable to find out, before method is to calculate the "power" of the study. This depends on five factors:

- the cutoff value (i.e. alpha level) below which the p-value from the study would be considered "statistically significant". This value is usually set at 0.05 or 5%;
- the disease rate in the non-exposed group in a cohort study or the exposure prevalence of the controls in a case-control study;

commencing the study, whether it is large enough to be informative. One

- the expected relative risk (i.e. the specified value of the relative risk under the alternative (non-null) hypothesis));
- the ratio of the sizes of the two groups being studied;
- the total number of study participants.

Once these quantities have been determined, standard formulas are then available to calculate the statistical power of a proposed study (Walter,

1977; Schlesselman, 1982). The standard normal deviate corresponding to the power of the study (derived from Rothman and Boice, 1982) is then:

$$Z_{\beta} = \frac{N_0^{0.5} |P_1 - P_0| B^{0.5} - Z_{\alpha} B}{K^{0.5}}$$

where:

- Z_{β} = standard normal deviate corresponding to a given statistical power
- Z_{α} = standard normal deviate corresponding to an alpha level (the largest p-value that would be considered "statistically significant")
- N_0 = number of persons in the reference group (i.e. the non-exposed group in a cohort study, or the controls in a case-control study)
- P_1 = outcome proportion in study group
- P_0 = outcome proportion in the reference group
- A = allocation ratio of referent to study group (i.e., the relative size of the two groups)
- $B = (1-P_0) (P_1 + (A-1) P_0) + P_0 (1-P_1)$
- $C = (1-P_0) (AP_1 - (A-1) P_0) + AP_0 (1-P_1)$
- $K = BC - A (P_1 - P_0)^2$

Standard calculator and microcomputer programmes incorporating procedures for power calculations are widely available. In particular, EPI-INFO (Dean et al, 1990) can be downloaded for free from

<http://www.cdc.gov/epiinfo/>, and Rothman's Episheet programme (Rothman, 2002) can be downloaded for free from <http://www.oup-usa.org/epi/rothman/>

Example 5.1

Consider a proposed study of 5,000 exposed persons and 5,000 non-exposed persons. Suppose that on the basis of mortality rates in a comparable

group of workers, the expected number of cases of the disease of interest is 25 in the non-exposed group. However, we expect that exposure will

double the risk of disease, so the number of cases observed will be 50 in the exposed group.

Then:

$$Z_{\alpha} = 1.96 \text{ (if a two-tailed significance test, for an alpha-level of 0.05, is to be used)}$$

$$N_0 = 5,000$$

$$P_1 = 0.010 \text{ (= 50/5000)}$$

$$P_0 = 0.005 \text{ (= 25/5000)}$$

$$A = 1$$

Using the equation above, the standard normal deviate corresponding to the power of the study to detect a

statistically significant lung cancer excess in the exposed group is:

$$Z_{\beta} = \frac{5000^{0.5} (0.010 - 0.005) (0.0149)^{0.5} - 1.96 \times 0.0149}{0.000197^{0.5}} = 0.994$$

From tables for the (one-sided) standard normal distribution, it can be seen that this corresponds to a power of 83%. This means that if 100 similar studies of this size were performed, then we would expect 83 of them to show a statistically significant ($p < 0.05$) excess of cases in the exposed group.

An alternative approach is to carry out a standard analysis of the hypothesized results. If we make the assumptions given above, then the relative risk would be 2.0, with a 90% confidence interval of 1.4-3.0. This approach only has an indirect relationship to the power calculations. For example, if the

lower 95% confidence limit is 1.0, then the power for a two-tailed test (of $p < 0.05$) would be only 50%. This simulated confidence interval gives the additional information that the observed relative risk could be as large as 3.0 or as low as 1.4 if the observed relative risk is 2.0.

Related approaches are to estimate the minimum sample sizes required to detect an association (e.g., relative risk) of specified magnitudes (Beaumont and Breslow, 1981), and to estimate the minimum detectable association for a given alpha level, power and study size (Armstrong, 1987).

Occasionally, the outcome is measured as a continuous rather than a dichotomous variable (e.g. blood pressure). In this situation the standard normal deviate corresponding to the study power is:

$$Z_{\beta} = \frac{N_0^{0.5}(\mu_1 - \mu_0) - Z_{\alpha}}{s(A + 1)^{0.5}}$$

where:

- μ_1 = mean outcome measure in exposed group
- μ_0 = mean outcome measure in reference group
- s = estimated standard deviation of outcome measure

The power is not the probability that the study will estimate the size of the association correctly. Rather, it is the probability that the study will yield a "statistically significant" finding when an association of the postulated size exists. The observed association could be greater or less than expected, but still be "statistically significant". The overemphasis on statistical significance is the source of many of the limitations of power calculations. Many features such as the significance level are completely arbitrary, issues of confounding, misclassification and effect modification are generally ignored (although appropriate methods are available - see Schlesselman,

1982; Greenland, 1983), and the size of the expected association is often just a guess. Nevertheless, power calculations are an essential aspect of planning a study since, despite all their assumptions and uncertainties, they nevertheless provide a useful general indication as to whether a proposed study will be large enough to satisfy the objectives of the study.

Estimating the expected precision can also be useful (Rothman and Greenland, 1998). This can be done by "inventing" the results, based on the same assumptions used in power calculations, and carrying out an analysis involving calculations of effect estimates and confidence limits. This approach has particular advantages when the exposure is expected to have no association with disease, since the concept of power is not applicable but precision is still of concern. However, this approach should be used with considerable caution, as the results may be misleading unless interpreted carefully. In particular, a study with an expected lower limit equal to a particular value (e.g. 1.0) will have only a 50% chance of yielding an observed lower confidence limit above that value.

In practice, the study size depends on the number of available participants and the available resources. Within these limitations it is desirable to make the study as large as possible, taking into account the trade-off between including more participants and gathering more detailed information about a smaller number of participants (Greenland, 1988). Hence, power calculations can only serve as a rough guide as to whether a feasible study is large enough to be worthwhile. Even if such calculations suggest that a particular study would have very low power, the study may still be worthwhile if exposure information is

collected in a form which will permit the study to contribute to the broader pool of information concerning a particular issue. For example, the International Agency for Research on Cancer (IARC) has organised several international collaborative studies such as those of occupational exposure to man-made mineral fibers (Simonato et al, 1986) and phenoxy herbicides and contaminants (Saracci et al, 1991). The man-made mineral fiber study involved pooling the findings from individual cohort studies of 13 European factories. Most of the

individual cohorts were too small to be informative in themselves, but each contributed to the overall pool of data.

Once a study has been completed, there is little value in retrospectively performing power calculations since the confidence limits of the observed measure of effect provide the best indication of the range of likely values for the true association (Smith and Bates, 1992; Goodman and Berlin, 1994). In the next chapter, random error will be ignored, and the discussion will concentrate on issues of systematic error.

Summary

Random error will occur in any epidemiologic study, just as it occurs in experimental studies. The most effective means of reducing random error is by increasing the study size, so that the precision of the effect estimate will be increased. Random error thus differs from systematic error which cannot be reduced simply by increasing the study size. The ideal study would be infinitely large, but practical considerations set limits on the number of participants that can be included. Given these limits, it is desirable to find out, before commencing the study, whether it is

large enough to be informative. One method is to calculate the "power" of the study. In practice, the study size depends on the number of available participants and the available resources. Within these limitations it is desirable to make the study as large as possible, taking into account the trade-off between including more participants and gathering more detailed information about a smaller number of participants. Hence, power calculations can only serve as a rough guide as to whether a feasible study is large enough to be worthwhile.

References

- Armitage P, Berry G, Matthews JNS (2002). *Statistical methods in medical research*. 4th ed. Oxford: Blackwell.
- Armstrong B (1987). A simple estimator of minimum detectable relative risk, sample size, or power in cohort studies. *Am J Epidemiol* 125: 356-358.
- Beaumont JJ, Breslow NE (1981). Power considerations in epidemiologic studies of vinyl chloride workers. *Am J Epidemiol* 114: 725-734
- Dean J, Dean A, Burton A, Dicker R (1990). *Epi Info*. Version 5.01. Atlanta, GA: CDC.
- Goodman SN, Berlin JA (1994). The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann Intern Med* 121: 200-6.
- Greenland S (1983). Tests for interaction in epidemiologic studies: a review and a study of power. *Statist Med* 2: 243-51.
- Greenland S (1988). Statistical uncertainty due to misclassification: implications for validation substudies. *J Clin Epidemiol* 41: 1167-74.
- Rothman KJ, Boice JD (1983). *Epidemiologic Analysis with a Programmable Calculator*. Epidemiology Resources, Inc.: Boston, MA.
- Rothman KJ (2002). *Epidemiology: an introduction*. New York, Oxford University Press.
- Rothman KJ, Greenland S (1998). *Modern epidemiology*. 2nd ed. Philadelphia: Lippincott-Raven.
- Saracci R, Kogevinas M, Bertazzi P, et al (1991). Cancer mortality in an international cohort of workers exposed to chlorophenoxy herbicides and chlorophenols. *Lancet* 338: 1027-32.
- Schlesselman JJ (1982). *Case-control studies: design, conduct, analysis*. New York: Oxford University Press.
- Simonato L, Fletcher AC, Cherrie J, et al. (1986). *Scandinavian Journal of Work, Environment and Health* 12: (suppl 1) 34-47.
- Smith AH, Bates M (1992). Confidence limit analyses should replace power calculations in the interpretation of epidemiologic studies. *Epidemiol* 3: 449-52.
- Walter SD (1977). Determination of significant relative risks and optimal sampling procedures in prospective and retrospective studies of various sizes. *Am J Epidemiol* 105: 387-97.

CHAPTER 6: Validity

(In: Pearce N. A Short Introduction to Epidemiology. 2nd ed. Wellington, CPHR, 2005)

Systematic error (lack of validity) is distinguished from random error (lack of precision) in that it would be present even with an infinitely large study, whereas random error can be reduced by increasing the study size. Thus, systematic error, or bias, occurs if there is a systematic difference between what the study is actually estimating and what it is intended to estimate.

There are many different types of bias, but in studies of cause and effect most biases fall into one of three major

categories (Rothman and Greenland, 1998): confounding; selection bias; and information bias. In general terms, these refer to biases arising from differences in baseline disease risk between the exposed and non-exposed subpopulations of the source population (confounding), biases resulting from the manner in which study participants are selected from the source population (selection bias), and biases resulting from the misclassification of these study participants with respect to exposure or disease (information bias).

6.1: Confounding

Confounding occurs when the exposed and non-exposed groups (in the source population) are not comparable due to inherent differences in background disease risk (Greenland and Robins, 1986) because of differences in the distribution of other risk factors between the exposed and non-exposed groups. For example, this could occur if we were studying the risk of heart disease in people who exercise frequently and those who do not, and if the people who exercised frequently smoked less than those who did not exercise; thus they might have a lower risk of heart disease because they smoked less, and not because they exercised more. Similar problems can occur in randomised trials because randomisation may fail, leaving the treatment groups with different

characteristics (and different baseline disease risk) at the time that they enter the study, and because of differential loss and non-compliance across treatment groups. However, there is more concern about non-comparability in epidemiological studies because of the absence of randomisation. The concept of confounding thus generally refers to the source population, although confounding can also be introduced (or removed) by the manner in which study participants are selected from the source population (Pearce and Greenland, 2004).

If no other biases are present, three conditions are necessary for a factor to be a confounder (Rothman and Greenland, 1998).

First, a confounder is a factor which is predictive of disease in the absence of the exposure under study. Note that a confounder need not be a genuine cause of disease, but merely "predictive". Hence, surrogates for causal factors (e.g. age) may be regarded as potential confounders, even though they are rarely directly causal factors.

Second, a confounder is associated with exposure in the source population at the start of follow-up (i.e. at baseline). In case-control studies this implies that a confounder will tend to be associated with exposure among the controls. An association can occur among the cases simply because the study factor and a potential confounder are both risk factors for the disease, but this does not cause confounding in itself unless the association also exists in the source population.

Thirdly, a variable which is affected by the exposure or the disease (e.g. an intermediate in the causal pathway

between exposure and disease, or a symptom of disease) should not be treated as a confounder because to do so could introduce serious bias into the results (Greenland and Neutra, 1981; Robins, 1987; Weinberg, 1993). For example, in a study of high fat diet and colon cancer, it would be inappropriate to control for serum cholesterol levels if it was considered that high serum cholesterol levels were a consequence of a high fat diet, and hence a part of the causal chain leading from diet to colon cancer. On the other hand, if serum cholesterol itself was of primary interest, then this should be studied directly, and high fat diet would be regarded as a potential confounder if it also involved exposure to other risk factors for colon cancer. Evaluating this type of possibility requires information external to the study to determine whether a factor is likely to be a part of the causal chain. Intermediate variables can sometimes be used in the analysis, but special techniques are then required to avoid adding bias (Robins, 1989; Robins et al, 1992; Robins et al, 2000).

Example 6.1

Table 6.1 presents a hypothetical example of confounding by tobacco smoking in a prevalence case-control study. One-half of the study participants are "exposed" to the risk factor of interest and one-half are not. However, two-thirds of the exposed people are smokers compared with one-third of the non-

exposed people. Thus, although "exposure" is not associated with disease either within the subgroup of smokers (POR=1.0) or within the subgroup of non-smokers (POR=1.0), it is associated with disease overall (POR=1.38) when the two subgroups are combined. This occurs because smoking is associated with the

exposure (as noted above) and is an independent risk factor for the disease (40% of non-exposed smokers have the disease compared with 20% of non-exposed non-smokers). Thus, smoking is a confounder and the "crude" prevalence odds ratio of 1.38 is invalid because it is not adjusted for smoking.

Table 6.1

Hypothetical example of confounding by tobacco smoking in a prevalence case-control study

	Smokers		Non-smokers		Total	
	Exposed	Non-exposed	Exposed	Non-exposed	Exposed	Non-exposed
Cases	800	400	200	400	1,000	8,00
Non-cases	1,200	600	800	1,600	2,000	2,200
Total	2,000	1,000	1,000	2,000	3,000	3,000
Prevalence (%)	40	40	20	20	33.3	26.7
Prevalence odds ratio	1.0		1.0		1.38	

Control of Confounding

Misclassification of a confounder leads to a loss of ability to control confounding, although control may still be useful provided that misclassification of the confounder is non-differential (Greenland, 1980). Misclassification of exposure poses a greater problem because factors which influence misclassification may appear to be confounders, but control of these factors may increase the net bias (Greenland and Robins, 1985). In general, control of confounding requires careful use of a priori knowledge, as well as inference from the observed data.

Control in the study design

Confounding can be controlled in the study design, or in the analysis, or both. Control at the design stage involves three main methods (Rothman and Greenland, 1998).

The first method is randomization, i.e., random allocation to exposure categories, but this is rarely an option in epidemiology which generally involves observational studies (it is debatable

whether randomised studies are part of epidemiology or whether they constitute a separate methodology).

A second method of control at the design stage is to restrict the study to narrow ranges of values of the potential confounders, e.g., by restricting the study to white males aged 35-54. This approach has a number of conceptual and computational advantages, but may severely restrict the number of potential study subjects and the generalizability of the study, as effects in younger or older people will not be observable.

A third method of control involves matching study subjects on potential confounders. For example, in a cohort study one would match a white male non-exposed subject aged 35-39 with an exposed white male aged 35-39. This will prevent age-sex-race confounding in a cohort study, but is seldom done because it may be very expensive. Matching can also be expensive in case-control studies, and does not prevent confounding in such studies, but does facilitate its control in the analysis. Matching may actually reduce precision in a case-control study if it is done on a

factor which is associated with exposure, but is not a risk factor for the disease of interest. However, matching on a strong risk factor will usually increase the precision of effect estimates.

Control in the Analysis

Confounding can also be controlled in the analysis, although it may be desirable to match on potential confounders in the design to optimize the efficiency of the analysis. The analysis ideally should control simultaneously for all confounding factors. Control of confounding in the analysis involves stratifying the data

according to the levels of the confounder(s) and calculating an effect estimate which summarizes the association across strata of the confounders. It is usually not possible to control simultaneously for more than 2 or 3 confounders in a stratified analysis. For example, in a cohort study, finer stratification will often lead to many strata containing no exposed or no non-exposed persons. Such strata are uninformative, thus fine stratification is wasteful of information. This problem can be mitigated to some extent, by the use of multiple regression which allows for simultaneous control of more confounders by "smoothing" the data across confounder strata.

Example 6.2

If the data presented in example 6.1 (table 6.1) is analysed separately in smokers and non-smokers, then the prevalence odds ratio is

1.0 in each of the two subgroups (i.e. 1.0 in smokers and 1.0 in non-smokers). Taking a weighted average of these two stratum-

specific estimates (see chapter 12) then yields an overall smoking-adjusted prevalence odds ratio of 1.0.

In general, control of confounding requires careful use of a priori knowledge, together with assessment of the extent to which the effect estimate changes when the factor is controlled in the analysis. Most epidemiologists prefer to make a decision based on the latter criterion, although it can be misleading, particularly if misclassification is present (Greenland and Robins, 1985). The decision to control for a presumed confounder can certainly be made with more confidence if there is supporting

prior knowledge that the factor is predictive of disease.

Misclassification of a confounder leads to a loss of ability to control confounding, although control may still be useful provided that misclassification of the confounder was nondifferential (unbiased) (Greenland, 1980). Misclassification of exposure is more problematic, since factors which influence misclassification may appear to be confounders, but control of these factors may increase the net bias (Greenland and Robins, 1985).

Example 6.3

Suppose that a cohort study of lung cancer involves a comparison with national mortality rates in a country where 50% of the population are non-smokers, 40% are moderate smokers with a 10-fold risk of lung cancer (compared to non-smokers), and 10% are heavy smokers with a 20-fold risk of lung cancer. Then, it can be calculated that the national lung cancer

incidence rate will be 6.5 ($= 0.50 \times 1.0 + 0.40 \times 10 + 0.10 \times 20$) times the rate in non-smokers. Suppose that it was considered most unlikely that the cohort under study contained more than 50% moderate smokers and 20% heavy smokers. Then, the incidence rate in the study cohort would be 9.4 times the rate in non-smokers. Hence, the observed incidence rate

would be biased upwards by a factor of $9.4/6.5 = 1.4$, i.e. it would be 1.4 times higher than the national rate due to confounding by smoking. Table 7.2 gives a range of such calculations presented by Axelson (1978) using data from Sweden. The last column indicates the likely bias in the observed rate ratio due to confounding by smoking (a value of 1.00 indicates no bias).

Table 6.2

Estimated crude rate ratios in relation to fraction of smokers in various hypothetical populations

	Population fraction (%)			Bias in relative risk
	Nonsmokers	Moderate Smokers ^a	Heavy Smokers ^a	
100	--	--	--	0.15
80		20	-	0.43
70		30	--	0.57
60		35	5	0.78
50		40	10	1.00 ^b
40		45	15	1.22
30		50	20	1.43
20		55	25	1.65
10		60	30	1.86
--		65	35	2.08
--		75	75	2.69
--		--	100	3.08

Source: Axelson (1978)

^aTwo different risk levels are assumed for smokers: 10 times for moderate smokers; and 20 times for heavy smokers.

^bReference population with rates similar to those in general population in countries such as Sweden.

Assessment of Confounding

When one lacks data on a suspected confounder (and thus cannot control confounding directly) it is still desirable to assess the likely direction and magnitude of the confounding it produces. It may be possible to obtain information on a surrogate for the confounder of interest (for example, social class is associated with many lifestyle factors such as smoking, and may therefore be a useful surrogate for some lifestyle-related confounders). Even though confounder control will be imperfect in this situation, it is still possible to examine whether the exposure effect estimate changes when the surrogate is controlled in the analysis, and to assess the strength and direction of the change. For example, if the relative risk actually increases (e.g. from 2.0 to 2.5), or remains stable (e.g. at 2.0) when social class is controlled for, then this is evidence that the observed excess risk is not due to confounding by smoking, since social class is correlated with smoking (Kogevinas et al, 1997), and control for social class involves partial control for smoking.

Alternatively, it may be possible to obtain accurate confounder information for a subgroup of participants in the study, and to assess the effects of confounder control in this subgroup. A related approach, known as two-stage sampling, involves obtaining confounder information for a sample of the source population (or a sample of the controls in a case-control study). For example, in a study of asthma in children, it may not be possible to obtain information on humidity levels in the home in all children. However, it may still be possible to obtain humidity measurements for a sample of the

exposed and non-exposed groups in order to check that the average level of humidity in the home is similar in the two groups. Such limited information, if taken in all exposure-disease subgroups, can also be used to directly control confounding (White, 1982; Walker, 1982; Rothman and Greenland, 1998).

Finally, even if it is not possible to obtain confounder information for any study participants, it may still be possible to estimate how strong the confounding is likely to be from particular risk factors. For example, this is often done in occupational studies, where tobacco smoking is a potential confounder, but smoking information is rarely available; in fact, although smoking is one of the strongest risk factors for lung cancer, with relative risks of 10 or 20, it appears that smoking rarely exerts a confounding effect of greater than 1.5 times in studies of occupational disease (Axelson, 1978; Siemiatycki, 1988), because few occupations are strongly associated with smoking, although this degree of confounding may still be important in some contexts.

6.2: Selection Bias

Whereas confounding generally involves biases that are inherent in the source population, and therefore would occur even if everyone in the source population took part in the study, selection bias involves biases arising from the procedures by which the study participants are selected from the source population. Thus, selection bias is not an issue in a cohort study involving complete follow-up, since in this case the study cohort composes the entire source population. However, selection bias can occur if participation

in the study or follow-up is incomplete. For example, in a cohort mortality study, if a national population registry (or some surrogate for this such as the United States Social Security system) were not available, then it might be necessary to attempt to contact each worker or his next-of-kin to verify vital status (i.e. whether the worker was still alive). Bias could occur if the response rate was higher in the most heavily exposed persons who had been diagnosed with disease than in other persons.

Example 6.4

Wrensch et al (2000) conducted a case-control study of 476 adults newly diagnosed with glioma in the San Francisco Bay Area between August 1991 and April 1994, and 462 age- gender- and ethnicity-matched controls. In addition, limited information was

obtained during a brief telephone interview with 101 controls who declined participation in the lengthy in-person interview. Controls who participated in the full interview were more likely than controls who only completed the telephone interview to report head injury. Thus

there was evidence of a selection bias in the recruitment of controls. The odds ratio for cases versus controls who completed the full interview was 0.9, whereas when both control groups were combined the odds ratio was 1.3.

Although we should recognize the possible biases arising from subject selection, it is important to note that epidemiologic studies need not be based on representative samples to avoid bias. For example, in a cohort study persons who develop disease might be more likely to be lost to follow-up than persons who did not develop disease; however, this would not affect the

relative risk estimate provided that loss to follow-up applied equally to the exposed and non-exposed populations (Criqui, 1979). Analogously, case-control studies have differing selection probabilities as an integral part of their design, in that the selection probability of diseased persons is usually close to 1.0 provided that most persons with disease are identified, whereas that for

non-diseased persons is substantially less; however, this does not affect the relative risk estimate provided that these selection probabilities apply equally within each exposure group.

Additional forms of selection bias can occur in case-control studies because these involve sampling from the source population. In particular, selection bias can occur in a case-control study (involving either incident or prevalent cases) if controls are chosen in a non-representative manner, e.g. if exposed people were more likely to be selected as controls than non-exposed people.

Minimizing Selection Bias

If selection bias has occurred in the enumeration of the exposed group, it may still be possible to avoid bias by choosing an appropriate non-exposed comparison group. For example, if the exposed group does not include all workers in a particular industry, but is

restricted to union members (because the records are available), then the non-exposed comparison group could be other workers in the same geographical area who are members of the same union, and/or a similar union.

Control of Selection Bias

Selection bias can sometimes be controlled in the analysis by identifying factors which are related to subject selection and controlling for them as confounders (provided that these factors are not affected by the study exposure or disease). For example, if white-collar workers are more likely to be selected for (or participate in) a study than manual workers (and white collar work is negatively or positively related to the exposure of interest), then this bias can be partially controlled by collecting information on social class and controlling for social class in the analysis as a confounder.

6.3: Information Bias

Information bias involves misclassification of the study participants with respect to disease or exposure status. Thus, the concept of information bias refers to those people actually included in the study, whereas selection bias refers to the selection of the study participants from the source population, and confounding generally refers to non comparability of subgroups within the source population. Information bias involves misclassification of the study subjects with respect to exposure, confounders, or disease.

It is customary to consider two types of misclassification: non-differential and differential misclassification.

Non-Differential Misclassification

Non-differential misclassification occurs when the probability of misclassification of exposure is the same for cases and non-cases (or when the probability of misclassification of disease is the same for exposed and non-exposed persons). This can occur if exposed and non-exposed persons are equally

likely to be misclassified according to disease outcome, or if diseased and non-diseased persons are equally likely to be misclassified according to exposure. Non-differential misclassification of exposure usually (but not always) biases the relative

risk estimate towards the null value of 1.0 (Copeland et al, 1977; Dosemeci et al, 1990). Hence, non-differential misclassification tends to produce "false negative" findings and is of particular concern in studies which find a negligible association

Example 6.5

In many cohort studies some exposed persons will be classified as non-exposed, and vice versa. Table 6.3 illustrates this situation with hypothetical data from a study of lung cancer incidence in asbestos workers. Suppose the true incidence rates are 100 per 100,000 person-years in the high exposure group, and 10 per 100,000 person-years in the low exposure group, and the relative

risk is thus 10. If 15% of high exposed persons are incorrectly classified, then 15 of every 100 deaths and 15,000 of every 100,000 person-years will be incorrectly allocated to the low exposure group. Similarly if 10% of high exposed persons are incorrectly classified, then 1 of every 10 deaths and 10,000 of every 100,000 person-years will be incorrectly allocated to the low exposure group. As a

result, the observed incidence rates per 100,000 person-years will be 91 and 23 respectively, and the observed relative risk will be 4.0 instead of 10.0. Due to non-differential misclassification, incidence rates in the high exposed group have been biased downwards, and incidence rates in the low exposure group have been biased upwards.

Table 6.3

Hypothetical data from a cohort study in which 15% of highly exposed persons and 10% of low exposed persons are incorrectly classified.

	Actual		Observed					
	High Exposure	Low Exposure	High Exposure		Low Exposure			
Deaths	100	10	85 +	1 =	86	9 +	15 =	24
Person-years	100,000	100,000	85,000 + 10,000 =		95,000	90,000 + 15,000 =		105,000
Incidence rate per 100,000 person years	100	10			91			23
Rate ratio	10.0						4.0	

between exposure and disease. One important condition is needed to ensure that exposure misclassification produces bias towards the null however: the exposure classification errors must be independent of other errors. Without this condition, non-differential exposure misclassification can produce bias in any direction (Chavance et al, 1992; Kristensen, 1992).

Furthermore, there are several other situations in which non-differential misclassification will not produce a bias towards the null.

Firstly, when the specificity of the method of identifying the disease under study is 100%, but the sensitivity is less than 100%, then the risk difference will be biased towards the null, but the risk ratio (or rate ratio) will be not be biased

by the misclassification. For example if only 80% of the deaths are identified in a study, but this under-ascertainment applies equally to the exposed and non-exposed groups, then this will not affect the relative risk estimate.

Secondly, the effect estimate may be biased away from the null for some exposure categories when there are multiple exposure categories (see example 6.6).

Finally, when there is positive confounding, and there is non-differential misclassification of the confounder, then confounding control will be incomplete and the adjusted effect estimate will consequently be biased away from the null.

Example 6.6

Table 6.4 gives hypothetical data from a cohort study in which the findings for the high and low exposure groups are the same as in example 6.5, but there is also a

non-exposed group for which there is no misclassification. In this instance, the non-differential misclassification between the high and low exposure

groups produces a bias away from the null when the low exposure group is compared to the non-exposed group: the relative risk is 4.6 instead of 2.0.

Table 6.4

Hypothetical data from a cohort study in which 15% of highly exposed persons and 10% of low exposed persons are incorrectly classified, but the non-exposed are correctly classified

	Actual			Observed		
	High	Low	Non-Exposed	High	Low	Non-Exposed
Deaths	100	10	5	86	24	5
Person-years	100,000	100,000	100,000	95,000	105,000	100,000
Rate	100	10	5	91	23	5
Rate ratio	20.0	2.0	1.0	18.1	4.6	1.0

One special type of non-differential misclassification occurs when the study outcome is not well-defined and includes a wide range of etiologically unrelated outcomes (e.g., all deaths). This may obscure the effect of exposure on one specific disease since a large increase in risk for this disease may only produce a small increase in risk for the overall group of diseases under study. A similar bias can occur when the exposure measure is not well defined and includes a wide range of etiologically unrelated exposures, possibly due to a non-specific exposure definition or due to the inclusion of exposures which could not have caused the disease of interest because they occurred after, or shortly before, diagnosis. It could be argued that these

phenomena do not represent misclassification because these are not errors in measurement. However, they do involve misclassification in the sense that the etiologically relevant exposure (or disease) has not been measured appropriately.

Differential Misclassification

Differential misclassification occurs when the probability of misclassification of exposure is different in diseased and non-diseased persons, or the probability of misclassification of disease is different in exposed and non-exposed persons. This can bias the observed effect estimate either toward or away from the null value. For example, in a nested case-control study of lung cancer, with a

Example 6.7

Table 6.5 shows data from a hypothetical case-control study in which 70 of the 100 cases and 50 of the 100 controls have actually been exposed to

some chemical. The true odds ratio is thus $(70/30) / (50/50) = 2.3$. If 90% (63) of the 70 exposed cases, but only 60% (30) of the 50 exposed controls

are classified correctly, then the observed odds ratio would be $(63/37) / (30/70) = 4.0$.

Table 6.5

Hypothetical data from a case-control study in which 90% of exposed cases and 60% of exposed controls are correctly classified

	Actual		Observed	
	Exposed	Non-exposed	Exposed	Non-exposed
Cases	70	30	63	37
Controls	50	50	30	70
Odds ratio	2.3		4.0	

control group selected from among non-diseased members of the cohort, the recall of occupational exposures in controls might be different from that of the cases. In this situation, differential misclassification would occur, and it could bias the odds ratio towards or away from the null, depending on whether members of the cohort who did not develop lung cancer were more or less likely to recall such exposure than the cases.

As can be noted from example 6.7, misclassification can drastically affect

the validity of a study. Given limited resources, it will often be more desirable to reduce information bias by obtaining more detailed information on a limited number of subjects than to reduce random error by including more subjects. However, a certain amount of misclassification is unavoidable, and it is usually desirable to ensure that it is towards the null value (as usually occurs with nondifferential exposure misclassification) to minimize the chance of false positive results.

Example 6.8

In the case-control study of lung cancer in Example 6.7, the misclassification could be made non-differential by selecting controls from cohort members with other types of cancer, or other diseases, in order that their recall of exposure would be more similar to that of the cases. As before, 63 (90%) of the

exposed cases would recall exposure, but now 45 (90%) of the 50 exposed controls would recall their exposure. The observed odds ratio would be $(63/37)/(45/55) = 2.1$. This estimate is still biased in comparison with the correct value of 2.3. However, the bias is non-differential, is much smaller than before, and

is in a predictable direction, towards the null. However, it should be noted that making a bias non-differential will not always make it smaller, and that the direction of bias from non-differential misclassification is sometimes predictable in advance.

Minimizing Information Bias

Misclassification can drastically affect the validity of a study. It is often helpful to ensure that the misclassification is non-differential, by ensuring that exposure information is collected in an identical manner in cases and non-cases (or that disease information is collected in an identical manner in the exposed and non-exposed groups). In this situation, if it is independent of other errors, exposure misclassification tends

to produce false negative findings and is thus of greatest concern in studies which have not found an important effect of exposure. Thus, in general it is important to ensure that information bias is non-differential and, within this constraint, to keep it as small as possible. Thus, can be argued that the aim of data collection is not to collect perfect information, but to collect information in a similar manner from

the groups being compared, even if this means ignoring more detailed exposure information if this is not available for both groups. However, this is not always the case (Greenland and Robins, 1985).

Assessment of information bias

Information bias is usually of most concern in historical cohort studies or case-control studies when information is obtained by personal interview. Despite these concerns, relatively little information is generally available on the accuracy of recall of exposures. When possible, it is important to attempt to validate the classification of exposure or disease, e.g., by comparing interview results with other data sources such as employer records, and to assess the potential magnitude of bias due to misclassification of exposure.

Relationship of Selection and Information Bias to Confounding

Selection bias and confounding are not always clearly demarcated. In particular, selection bias can sometimes be viewed as a type of confounding, since both can be reduced by controlling for surrogates for the determinants of the bias (e.g. social class). Unfortunately, selection affected by exposure and disease generates a bias that cannot be reduced in this fashion. Some consider any bias that can be controlled in the analysis as confounding. Other biases are then categorized according to whether they arise from the selection of study subjects (selection bias), or their classification (information bias).

Summary

The greatest concern in epidemiological studies usually relates to confounding, because exposure has not been randomly allocated, and the groups under study may therefore be noncomparable with respect to their baseline disease risk. However, to be a significant confounder, a factor must be strongly predictive of disease and strongly associated with exposure. Thus, although confounding is constantly a source of concern, the strength of confounding is often considerably less than might be expected (it should be appreciated however, that this appearance may be illusory, for nondifferential misclassification of a confounder which is common will usually make the

confounding appear smaller than it really is).

Provided that information has been collected in a standardized manner (and its accuracy is unrelated to other errors), then misclassification will be non-differential, and any bias it produces will usually be towards the null value. In this situation, misclassification tends to produce false negative findings and is thus of greatest concern in studies which have not found an important effect of exposure; it is of much less concern in studies with positive findings, since these findings are likely to have been even more strongly positive if misclassification had not occurred.

Again, one should appreciate the limitations of these observations: it may be difficult to be sure that the exposure and disease misclassification is nondifferential, and nondifferential

misclassification of a confounder can lead to bias away from the null if the confounder produces confounding away from the null.

References

- Axelsson O (1978). Aspects on confounding in occupational health epidemiology. *Scand J Work Environ Health* 4: 85-9.
- Chavance M, Dellatolas G, Lellouch J (1992). Correlated nondifferential misclassifications of disease and exposure: application to a cross-sectional study of the relationship between handedness and immune disorders. *Int J Epidemiol* 21: 537-46.
- Copeland KT, Checkoway H, McMichael AJ, et al (1977). Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol* 105: 488-95.
- Criqui MH (1979). Response bias and risk ratios in epidemiologic studies. *American Journal of Epidemiology* 109:394-399.
- Dosemeci M, Wacholder S, Lubin JH (1990). Does nondifferential misclassification of exposure always bias a true effect toward the null value? *Am J Epidemiol* 132: 746-8.
- Greenland S (1980). The effect of misclassification in the presence of covariates. *Am J Epidemiol* 112: 564-9.
- Greenland S, Neutra R (1981). An analysis of detection bias and proposed corrections in the study of estrogens and endometrial cancer. *J Chron Dis* 34: 433-8.
- Greenland S, Robins JM (1985). Confounding and misclassification. *Am J Epidemiol* 122: 495-506.
- Greenland S, Robins JM (1986). Identifiability, exchangeability and epidemiological confounding. *Int J Epidemiol* 15: 412-8.
- Kogevinas M, Pearce N, Susser M, Boffetta P (1997). Social inequalities and cancer. In: Kogevinas M, Pearce N, Susser M, Boffetta P (eds). *Social inequalities and cancer*. Lyon: IARC, pp 1-15.
- Kristensen P (1992). Bias from nondifferential but dependent misclassification of exposure and outcome. *Epidemiol* 3: 210-5.
- Pearce N, Greenland S (2004). Confounding and interaction. In: Ahrens W, Krickeberg K, Pigeot I (eds). *Handbook of epidemiology*. Heidelberg: Springer-Verlag, 2004, pp 375-401.
- Robins J (1987). A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *J Chron Dis* 40 (suppl 2): 139S-161S.

Robins J (1989). The control of confounding by intermediate variables. *Stat Med* 8: 679-701.

Robins JM, Blevins D, Ritter G, et al (1992). G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of AIDS patients. *Epidemiol* 3: 319-36.

Robins JM, Hernán MA, Brumback B (2000). Marginal structural models and causal inference in epidemiology. *Epidemiol* 11; 550-62.

Rothman KJ, Greenland S (1998). *Modern epidemiology*. 2nd ed. Philadelphia: Lippincott-Raven.

Siemiatycki J, Wacholder S, Dewar R, et al (1988). Smoking and degree of occupational exposure: Are internal analyses in cohort studies likely to be confounded by smoking status?

Walker AM (1982). Anamorphic analysis: sampling and estimation for covariate effects when both exposure and disease are known. *Biometrics* 38: 1025-32.

Weinberg CR (1993). Toward a clearer definition of confounding. *Am J Epidemiol* 137: 1-8.

White JE (1982). A two-stage design for the study of the relationship between a rare exposure and a rare disease. *Am J Epidemiol* 115: 119-28.

Wrensch M, Miike R, Neuhaus J (2000). Are prior head injuries of diagnostic X-rays associated with glioma in adults? The effects of control selection bias. *Neuroepidemiology* 2000; 19: 234-44.

CHAPTER 7: Effect Modification

(In: Pearce N. A Short Introduction to Epidemiology. 2nd ed. Wellington, CPHR, 2005)

In the previous chapter I discussed the problem of confounding which occurs when the exposed and non-exposed subpopulations of the source population are inherently different in background disease risk. This should not be confused with effect modification which occurs when the measure of the effect of the study factor depends on the level of another factor in the study population

(Miettinen, 1974). The term statistical interaction denotes a similar phenomenon in the observed data. However, the terms "interaction" and "effect modification" are also used in a variety of other contexts, with a variety of meanings. In particular, the term "interaction" has different meanings for biostatisticians, lawyers, clinicians, public health professionals, epidemiologists and biologists.

Example 7.1

Katsouyanni et al (1993) studied the effects of air pollution and high temperature in the causation of excess mortality during a major heat wave in Greece in July 1987. They found that the effects of the

heat wave were modified by the presence (or absence) or high air pollution levels. In Athens (where air pollution levels are high) the increase in deaths on extremely hot days was 97% in Athens, but was 33% in other

urban areas and 27% in non-urban areas. Further analyses suggested that the threshold of effect of various air pollutants appeared to be lower on extremely hot days.

7.1: Concepts of Interaction

The different concepts of interaction will be illustrated with data from a hypothetical study of the risk of lung cancer per 1,000 population (e.g. over a five year period) in relation to exposure to cigarette smoke and asbestos (Table 7.1). The risk difference due to smoking is 30 per 1,000 in asbestos workers and 9

per 1,000 person-years in smokers. On the other hand, the rate ratio for smoking is 7.0 in asbestos workers and 10.0 in other people. I will now consider how this data might be interpreted by a different researchers and policy makers. In each instance, it is recognized that it is important to prevent or reduce both

asbestos exposure and smoking. However, in this case the asbestos exposure has already occurred and the factory has now closed, so our focus is on smoking. We want to know whether the 'effect' of smoking is "modified" by asbestos exposure, i.e. do smoking and asbestos exposure "interact"?

Two Biostatisticians

Suppose that we first consult a biostatistician about how to interpret this data. The first biostatistician we talk to uses relative risk measures of effect. They note that the relative risk for smoking and lung cancer is 7.0 (35/5) in asbestos workers and 10.0 (10/1) in other people. Thus, the effect of smoking on lung cancer is less in asbestos workers and there is therefore a negative statistical interaction between the effects of smoking and asbestos (table 7.2). They may even fit a multiplicative model with an interaction term and show that the interaction term is negative.

We can see the logic of this argument, but are somewhat surprised by the conclusion, since we can see the very high rates in people who both smoke and are exposed to asbestos. We therefore consult a second biostatistician. This "alternative" biostatistician uses the risk

difference as the effect measure. They note that the risk difference for smoking and lung cancer is 30 per 1,000 (35 - 5) in asbestos workers and 9 per 1,000 (10 - 1) in other people. Thus, the effect of smoking is greater in asbestos workers and there is a positive statistical interaction between the effects of smoking and asbestos (table 7.2). They may even fit an additive model with an interaction term and show that the interaction term is positive.

We eventually get our two biostatistical consultants together and they argue that there is no contradiction in the advice they have given us. Effect modification and statistical interaction are merely statistical concepts which depend on the methods used. In fact, all secondary risk factors modify either the rate ratio or the rate difference, and uniformity over one measure implies non-uniformity over the other (Koopman, 1981; Steenland and Thun, 1986), e.g. an apparent additive joint effect implies a departure from a multiplicative model. Several authors (e.g. Kupper and Hogan, 1978; Walter and Holford, 1978) have demonstrated the dependence of statistical interaction on the underlying statistical measure of effect, and have therefore argued that the assessment of interaction is "model-dependent".

Table 7.1

Lung cancer risk per 1,000 people (and RR) in relation to exposure to cigarette smoke and asbestos

	Asbestos	
	Yes	No
Smoking	Yes	35/1000 (35.0)
	No	5/1000 (5.0)
Rate difference	30/1000	9/1000
Rate ratio	7.0	10.0

A Lawyer

Next we consult a lawyer (I do not advise this as a real course of action; this is just a hypothetical consultation!). She/he is also concerned about the effect of smoking, but the effect they are interested in is "what is the probability that my client's lung cancer was caused by their smoking?" If we look at the asbestos workers, we find that if they smoked their risk of lung cancer was 35 per 1,000 whereas it was 7 per 1,000 if they didn't smoke. Thus, assuming there is no confounding by other factors, then of every 35 lung cancer occurring in the smokers, 5 would have happened anyway, and 30 are additional cases due to smoking. Thus, for an individual lung cancer case, the probability that smoking caused the cancer is $100 \times 30/35$ which is

86% (this is just $100 \times (R-1)/R$ where R is the relative risk of 1.9). The corresponding estimate for other people (not exposed to asbestos) is $100 \times 9/10$ which is 90%. Thus, the probability of causation by smoking is slightly less in asbestos workers and there is therefore a negative interaction between the effects of smoking and asbestos (table 7.2). It should be noted that this lawyer's approach is a little simplistic (Greenland, 1999), but the key issue here is that the "effect" that is being measured, and the inference about interaction, is different from that of the two biostatisticians, although it is more consistent with that of the biostatistician who uses the relative risk as the measure of effect.

Table 7.2

The approaches of different consultants to interpreting the data in table 8.1

Consultant	Effect measure	Size of effect		Inherent Statistical model	Is there an Interaction? Direction?	
		Asbestos workers	Others			
Biostatistician 1	Relative risk	7.0	10.0	Relative risk	Yes	-ve
Biostatistician 2	Risk difference	30/1000	9/1000	Risk difference	Yes	+ve
Lawyer	Probability of causation	86%	90%	Relative risk	Yes	-ve
Clinician	Individual risk	30 per 1,000	9 per 1,000	Risk difference	Yes	+ve
Public health worker	Deaths prevented	30 per 1,000	9 per 1,000	Risk difference	Yes	+ve
Epidemiologist	Combination of factors to cause disease	21 cases out of 35 (60%) are due to the combination of exposures	Not applicable	Risk difference	Yes	+ve

A Clinician

Next we consult with a clinician. She/he says "I advise my patients to give up smoking, and I tell them that if they do manage to stop then they will reduce their risk of lung cancer. They ask 'by how much?' So I want to know what the reduction in their individual risk will be if they give up smoking". Well, if their patient is an asbestos worker then they will reduce their risk by 30 per 1,000 (over five years) by giving up smoking; other people will reduce their risk by 9 per 1,000 (once again, this is a little simplistic since it does not tell us exactly how many years of life they will gain). Thus, the effect of smoking is greater in asbestos workers and there is therefore a positive statistical interaction between the effects of smoking and asbestos (table 8.2).

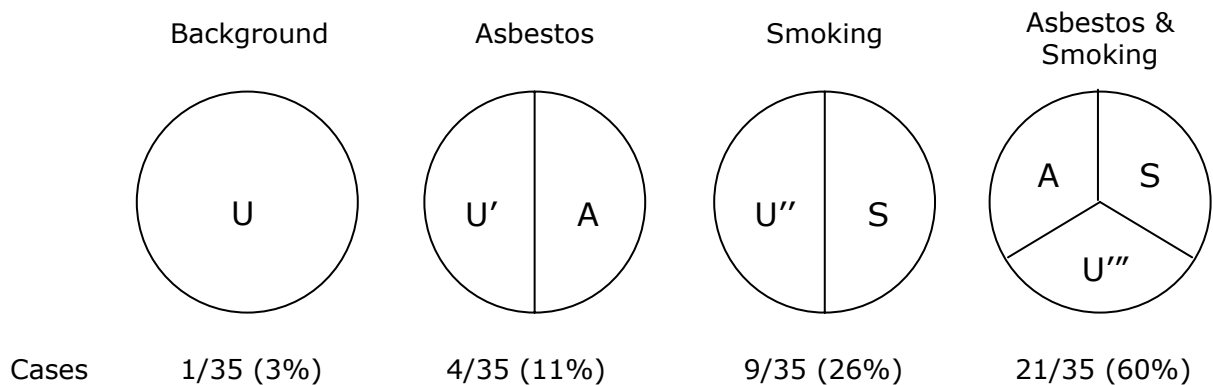
A Public Health Worker

The public health worker that we consult has a similar approach to the clinician,

except that they are concerned about the population rather than about individual patients. They say "I want to conduct population smoking prevention campaigns and persuade people to give up smoking and that if they do then they will reduce their risk of lung cancer. I only have a limited amount of resources so I want to know if I can prevent more cases of lung cancer by focusing on asbestos workers, or by doing my campaigns in the same number of people in the general population". If they prevent 1,000 asbestos workers smoking, then (once there has been time for the reduction in risk to start occurring) they will have prevented 30 lung cancer cases each year. If they prevent 1,000 other people from smoking then each year they will have prevented 9 cases of lung cancer. Thus, the effect of smoking is greater in asbestos workers and there is therefore a positive statistical interaction between the effects of smoking and asbestos (table 7.2).

Figure 7.1

Numbers of cases occurring through background factors, asbestos alone, smoking alone, and their combination in people exposed to both factors



An Epidemiologist

I have argued in chapter 1 that epidemiology is part of public health, and therefore I might be quite content to accept the public health worker's approach. However, as an epidemiologist I do want to know more about the causation of disease, since what I learn may be relevant to other exposures or other diseases. Thus, I may be particularly interested in the combination of smoking and asbestos to produce cases of lung cancer. Rothman and Greenland (1998) have thus adopted an unambiguous epidemiological definition of interaction in which two factors are not "independent" if they are component causes in the same sufficient cause. This concept of independence of effects leads to the adoption of additivity of incidence rates as the state of "no interaction". Thus, the fact that the lung cancer rate in the group exposed to both factors (35/1000) is greater than the sum of the baseline risk (1/1000) plus the effect of asbestos alone (5/1000 - 1/1000) plus the effect of smoking alone (10/1000 - 1) indicates that there are some cases of disease that are occurring due to the combination of exposures and which would not have occurred if either of the exposures had been eliminated. We can do the same calculations using the relative risks (relative to the group with exposure to neither factor) rather than incidence rates: the joint effect is 35.0 times, whereas it would be $1+(5.0-1)+(10.0-1)=14.0$ if it were additive. This situation is summarized in figure 7.1. It shows that in the group exposed to both factors, 1 case (3%) occurred through unknown "background" exposures (U), 4 cases (11%) through mechanisms involving asbestos exposure alone (and not smoking) together with unknown background exposures (U'), 9 cases (26%) occurred through mechanisms involving smoking alone (and not

asbestos) together with unknown background exposures (U''), and 21 cases (60%) occurred through mechanisms involving both factors together with unknown background exposures (U'''). This means that 86% of the cases (26% + 60%) could have been prevented by preventing smoking, whereas 71% (11% + 60%) could have been prevented by preventing asbestos exposure. Thus, the attributable risks for the individual factors of smoking (86%) and asbestos (71%) sum to more than 100% because of the cases that occur through mechanisms involving both exposures and which consequently could be prevented by preventing either exposure.

One apparent exception should be noted (Koopman, 1977). If two factors (A and B) belong to different sufficient causes, but a third factor (C) belongs to both sufficient causes, then A and B are competing for a single pool of susceptible individuals (those who have C). Consequently the joint effect of A and B will be less than additive (Miettinen (1982) reaches a similar conclusion based on a model of individual outcomes). However, this phenomenon can be incorporated directly into the causal constellation model by clarifying a previous ambiguity in the description of antagonism in the model's terms. Specifically, the absence of B can be included in the causal constellation involving A, and vice versa. Then, two factors would not be "independent" if the presence or absence of the factors (or particular levels of both factors) were component causes in the same sufficient cause (Greenland and Poole, 1988; Rothman and Greenland, 1998).

A Biologist

Finally, it should be stressed that this epidemiological concept of independence of effects is distinct from

some biological concepts of independence. For example, Siemiatycki and Thomas (1981) give a definition in which two factors are considered to be biologically independent "if the qualitative nature of the mechanism of action of each is not affected by the presence or absence of the other". However, this concept does not lead to an unambiguous definition of independence of effects, and thus does not produce clear analytic implications. Rothman's concept of independence is at a more abstract conceptual level in

which a particular biologic model, rather than being accepted as the "baseline", is itself evaluated in terms of the co-participation of factors in a sufficient cause. For example, two factors which act at different stages of a multistage process are not independent since they are joint components of at least one sufficient cause. This occurs irrespective of whether they affect each other's qualitative mechanism of action (the ambiguity in Siemiatycki and Thomas' formulation stems from the ambiguity of this concept).

7.2 Multiplicative and Additive Models

Rothman's approach is attractive because it is based on epidemiological concepts which have a clear biologic interpretation, and because it leads to an unambiguous definition of independence of effects which is identical to that obtained through public health considerations (Rothman et al, 1980). However, the analytic implications of these concepts are not straightforward, since assessing independence of effects is usually only one of the analytic goals of an epidemiological study. Rather, there are several other considerations which often favour the use of multiplicative models.

First, multiplicative models have convenient statistical properties. Estimation in non-multiplicative models may have problems of convergence, and inference based on the asymptotic standard errors may be flawed unless the study size is very large (Moolgavkar and Venzon, 1987).

Second, it has been argued that multiplicative models facilitate the assessment of the extent of unknown confounding or bias (Cornfield et al, 1959), although this is not always the case.

Third, if it is desired to keep statistical interaction (effect modification) to a minimum, then a multiplicative model may be more appropriate. It is not uncommon for risk factors to have approximately multiplicative effects (Saracci, 1987). This presumably occurs because they are a part of common causal processes, although other sufficient causes usually also operate, and exact multiplicativity may not occur. Nevertheless, in this situation there may be less masking of heterogeneity in calculating an overall rate ratio than in calculating an overall rate difference; there are also many instances of non-multiplicative departures from additivity, however (Selikoff et al, 1980; Saracci, 1987).

7.3: Joint Effects

These considerations imply an apparent dilemma. How can an analysis be conducted which combines the advantages of ratio measures of effect with the assessment of independence in terms of a departure from additivity? These apparently contradictory goals can be reconciled in analyses which concentrate on the estimation of separate and joint effects (Pearce, 1989).

Thus, when studying asbestos, smoking and lung cancer, relative risks might be presented for smoking (in non-asbestos workers), asbestos exposure (in non-smokers) and exposure to both factors, relative to persons exposed to neither factor. These relative risks would be adjusted for all other factors (e.g. age) which are potential confounders, but not of immediate interest as effect modifiers.

The estimation of separate and joint effects may be difficult when the factors of interest are closely correlated, and there are therefore only small numbers of people who are exposed to either factor alone. However, when it is feasible, this approach combines the best features of multiplicative models and additive independence assessment, but also permits readers with other concepts of independence to draw their own conclusions (as in table 7.1).

When the assessment of joint effects is a fundamental goal of the study, it can be accomplished by calculating stratum-specific effect estimates, as in Example 7.1 above. On the other hand, it is less clear how to proceed when effect modification is occurring, but assessment of joint effects is not an analytical goal. Conventional statistical

analysis strategies are based on the principle that it is not appropriate to calculate an overall effect estimate if interaction is present. However, this principle is commonly ignored if the difference in stratum-specific effect estimates is not too great. In fact standardized rate ratios (see chapter 12) have been developed for precisely this situation, and will consistently estimate meaningful epidemiological parameters even under heterogeneity (Greenland, 1982). Nevertheless, some authors have proposed modeling strategies in which the first step in the analysis involves testing for statistical interaction. A related approach has been the development of generalized families of models which include the additive and multiplicative models as special cases. An alternative general strategy can be based on epidemiological considerations (Pearce, 1989). The key difference is that interaction is assessed (rather than tested) in terms of a departure from additivity in order to elaborate an observed effect, rather than being tested for departure from an arbitrary effect measure as an essential initial analytic step. This procedure can be achieved within the confines of statistically convenient multiplicative models through the analysis of separate and joint effects.

Summary

The terms interaction and effect modification are used in a variety of contexts, with a variety of meanings. In particular, the term "interaction" has different meanings for biostatisticians, lawyers, clinicians, public health professionals, epidemiologists and biologists. In each instance, they are interested in the same question, namely does the effect of exposure A depend on whether exposure B is also present (or absent)? However, the word "effect" has different meanings in different contexts. In contrast to definitions based on statistical concepts, Rothman has adopted an unambiguous epidemiological definition of interaction

in which two factors are not "independent" if they are component causes in the same sufficient cause. This leads to the adoption of additivity of incidence rates as the state of "no interaction". However, there are other considerations which generally favor the use of multiplicative models. This implies an apparent dilemma as to how an analysis can be conducted which combines the advantages of ratio measures of effect with the assessment of independence in terms of a departure from additivity. These apparently contradictory goals can be reconciled through the analysis of separate and joint effects.

References

- Cornfield J, Haenszel W, Hammond EC, et al (1959). Smoking and lung cancer: recent evidence and a discussion of some questions. *JNCI* 22: 173-203.
- Greenland S (1982). Interpretation and estimation of summary ratios under heterogeneity. *Statist Med* 1: 217-27.
- Greenland S, Poole C (1988). Invariants and noninvariants in the concept of interdependent effects. *Scand J Work Environ Health* 14; 125-9.
- Greenland S (1999). Relation of probability of causation to relative risk and doubling dose: a methodologic error that has become a social problem. *AJPH* 89: 1166-9.
- Katsouyanni K, Pantazopoulou A, Touloumi G, et al. Evidence for interaction between air pollution and high temperature in the causation of excess mortality. *Arch Environ Health* 1993; 48: 235-42.
- Koopman JS (1977). Causal models and sources of interaction. *Am J Epidemiol* 106: 439-44.
- Koopman JS (1981). Interaction between discrete causes. *American Journal of Epidemiology* 113:716-724.
- Kupper LL, Hogan MD (1978). Interaction in epidemiologic studies. *Am J Epidemiol* 108: 447-53.
- Miettinen OS (1974). Confounding and effect modification. *Am J Epidemiol* 100: 350-3.

- Miettinen OS (1982). Causal and preventive interdependence: elementary principles. *Scand J Work Environ Health* 8: 159-68.
- Moolgavkar SH, Venzon DJ (1987). General relative risk models for epidemiologic studies. *Am J Epidemiol* 126: 949-61.
- Pearce NE (1989). Analytic implications of epidemiological concepts of interaction. *Int J Epidemiol* 18: 976-80.
- Rothman KJ, Greenland S, Walker AM (1980). Concepts of interaction. *Am J Epidemiol* 112: 467-70.
- Rothman KJ, Greenland S (1998). *Modern epidemiology*. 2nd ed. Philadelphia: Lippincott-Raven.
- Saracci R (1987). The interactions of tobacco smoking and other agents in cancer etiology. *Epidemiologic Reviews* 9: 175-93.
- Selikoff I, Sedman H, Hammond E (1980). Mortality effects of cigarette smoking among amosite asbestos factory workers. *JNCI* 65: 507-13
- Siemiatycki J, Thomas DC (1981). Biological models and statistical interactions: an example from multistage carcinogenesis. *Int J Epidemiol* 10: 383-7.
- Steenland K, Thun M (1986). Interaction between tobacco smoking and occupational exposures in the causation of lung cancer. *Journal of Occupational Medicine* 28:110-118.
- Walter SD, Holford TR (1978). Additive, multiplicative and other models for disease risks. *Am J Epidemiol* 108: 341-6.

Part III

Conducting a study

CHAPTER 8: Measurement of Exposure and Health Status

(In: Pearce N. A Short Introduction to Epidemiology. 2nd ed. Wellington, CPHR, 2005)

In this chapter I briefly review the various options for measuring exposure and disease status. In the following

chapters I then discuss the practicalities of conducting cohort, case-control and cross-sectional studies.

8.1: Exposure

As discussed in chapter 1, epidemiological studies involve a wide variety of exposures ranging from the population level to the individual and micro-levels. The term "exposure" is thus used generically to refer to any factor that is under study, and exposures may include population factors (e.g. income inequality), individual-level socio-economic factors (e.g. income), physical environmental factors (e.g. air pollution), aspects of individual lifestyle (e.g. diet), as well as "exposures" measured at the level of the body, (e.g. total body burden of dioxin), organ (e.g. the concentration of asbestos in the lung), cell, or molecule (e.g. DNA adducts). These various situations are discussed here briefly; a more detailed discussion can be found in Armstrong et al (1992).

Exposure and Dose

Strictly speaking, the term exposure refers to the presence of a substance (e.g. fine particulate matter) in the external environment, whereas the term dose refers to the amount of substance that reaches susceptible targets within the body, such as the airways. In some situations (e.g. in a coal mine) measurements of external exposures

may be strongly correlated with internal dose, whereas in other situations (e.g. environmental lead exposure) the dose may depend on individual lifestyle and activities and may therefore be only weakly correlated with the environmental exposure levels.

Exposure levels can be assessed with regard to the intensity of the substance in the environment (e.g. dust concentration in the air) and the duration of time for which exposure occurs. The risk of developing disease may be much greater if the duration of exposure is long and/or the exposure is intense, and the total cumulative exposure may therefore be important. For protracted etiologic processes, the time-pattern of exposure may be important and it is possible to assess this by examining the separate effects of exposures in various time windows prior to the occurrence and recognition of clinical disease (Pearce, 1992). For example, in cancer studies recent exposures may not be relevant since the cancer may have first become established some years previously (Pearce, 1988). Similarly, recent work suggests that occupational asthma is most likely to occur after about 1-3

years of exposure to a sensitising agent (Antó et al, 1996).

General Approaches to Exposure Assessment

Methods of exposure measurement include personal interviews or self-administered questionnaires (completed either by the study participant or by a proxy respondent), diaries, observation, routine records, physical or chemical measurements on the environment, or physical or chemical measurements on the person (Armstrong et al, 1992). For example, table 8.1 summarizes the types of exposures data most commonly used in occupational epidemiology studies (Checkoway et al, 2004). Measurements on the person can relate either to exogenous exposure (e.g. airborne dust) or internal dose (e.g. plasma cotinine); the other measurement options (e.g. questionnaires) all relate to exogenous exposures.

Demographic Factors

In most instances, information on demographic factors such as age, gender and ethnicity can be obtained in a straightforward manner from routine health care records or with questionnaires. In studies focusing on ethnicity, the etiologically relevant definition will depend on the extent to which an ethnic difference is considered to be due to genetic and/or cultural and environmental factors, but the available information will vary from country to country depending on historical and cultural considerations. For example, in New Zealand, Māori ethnicity is defined as 'a person who has Māori ethnicity and chooses to identify as Māori' (Pomare et al, 1992), whereas some other countries use solely biologically-based definitions of "race" or ethnicity (Polednak, 1989).

Socio-economic status poses more significant measurement problems. It

can be measured in a variety of ways, including occupation, income, and education (Liberatos et al, 1988; Berkman and MacIntyre, 1997). These measures may pose problems in some demographic groups; for example, occupation and income may be poor measures of socio-economic status in women, for whom the total family situation may reflect their socio-economic status better than their individual situation, and measures of socio-economic status in children must be based on the situation of the parents or the total family situation.

Nevertheless, the various measures of socio-economic status are strongly correlated with each other, and asthma epidemiology studies are usually based on whichever measures are available, unless socio-economic status is the main focus of the research and it is necessary to obtain more detailed information.

Questionnaires

Traditionally, exposure to most non-biological risk factors (e.g. tobacco smoking) has been measured with questionnaires, and this approach has a long history of successful use in epidemiology (Armstrong et al, 1992). Questionnaires may be self-administered (e.g. postal questionnaires) or interviewer-administered (e.g. in telephone or face-to-face interviews) and may be completed by the study subject or by a proxy (e.g. parental completion of questionnaires in a study of children, or completion by the spouse of deceased cases). The validity of questionnaire data also depends on the structure, format, content and wording of questionnaires, as well as methods of administration and selection and training of interviewers (Armstrong et al, 1992).

Table 8.1

Types of exposure data commonly used in occupational epidemiology studies
(Source: Adapted from Checkoway et al, 2004)

- Ever employed in the industry
- Duration of employment in the industry
- Ordinally ranked jobs or tasks
- Job-exposure matrices
- Quantified personal measurements

Example 8.1

Raum et al (2001) studied the impact of maternal socio-economic status on intrauterine growth in the former west and East Germany. Information on socio-demographic or lifestyle factors and pregnancy outcome was available for 3,374 live-born singletons from West Germany (1987/88) and 3070 from East Germany (1990/91). Women were recruited during

pregnancy and given a self-administered 30-page questionnaire covering socio-demographic, psychosocial, nutritional, environmental and occupational factors. The two school systems were not identical, but in each system maternal educational level was grouped into five categories. Women with the lowest education had a significantly elevated

risk of small-for-gestational-age (SGA) newborns compared to women with the highest education in both the west (OR = 2.58, 95% CI 1.17-5.67) and the east (OR = 2.77, 95% CI 1.54-5.00). The authors concluded that social inequalities existed and caused health inequalities in both the West, and in the former socialist country of East Germany.

Example 8.2

Vartia (2001) studied the consequences of workplace bullying in the municipal sector in Helsinki, Finland. Every 35th member of the Municipal Officials Union was selected and 1037 (65.5%) responded to a postal questionnaire. A definition of bullying was provided and study

participants were asked if they felt themselves subjected to such behaviour, or if they had observed someone else at their workplace being bullied. They were also asked about the frequency and duration of such acts. Both the targets of bullying and the observers reported

more general stress and mental stress reactions than did respondents from workplaces with no bullying. The targets of bullying used sleep-inducing drugs and sedatives more often than did the respondents who were not bullied.

Environmental Measurements and Job-Exposure Matrices

In many studies, e.g. community-based case-control studies, questionnaires are the only source of exposure information. However, in some instances, particularly in occupational studies, questionnaires may be combined with environmental exposure measurements (e.g. industrial hygiene surveys) to obtain a quantitative estimate of individual exposures. Table 8.2 shows environmental measurements in an asbestos textile plant in South Carolina (Dement et al, 1983; Checkoway et al, 2004). It shows that

for each job title, exposure levels decreased over time, but increased again during the 1966-75 time period. Within each time period, the highest exposures were in raw fiber handling and the lowest were in general area workers. This historical exposure information can be combined with information from employment records to obtain exposure estimates for individual workers. For example, table 5.3 shows the cumulative exposure for a worker who worked as a card operator during 1933-1938 and then worked in "clean-up" during 1939-1948.

Example 8.3

Saracci et al (1984a) conducted a historical cohort study of mortality and cancer incidence of workers exposed to made-made vitreous fibres at 13 European plants. At 12 of the plants an environmental survey was conducted to measure present concentrations of fibres in air samples. This was used to create a job-exposure matrix. Within

each plant, job/plant areas were grouped into six main occupational categories: not specified, office, preproduction, production, secondary processes and maintenance. For each worker a cumulative exposure index was created by multiplying the time spent in each job category by the mean concentration of

respirable fibres in the job category. The relative risk of lung cancer was elevated, particularly in the group with 30 years or more since first employment (RR=1.92, 95% CI 1.17-3.07). There was a tendency for the risk to increase with cumulative exposure, but the pattern was not consistent.

Table 8.2

Asbestos concentrations (fibers/cc) in job categories in an asbestos textile plant
(Source: Adapted from Checkoway et al, 2004)

Job category	1930-35	1936-45	1946-65	1966-75
General area	10.8	5.3	2.4	4.3
Card operators	13.3	6.5	2.9	5.3
Clean-up	18.1	8.8	4.0	7.2
Raw fiber handling	22.8	11.0	5.0	9.0

Table 8.3

Example of an exposure history of an individual worker

Job	Years	Mean exposure	Cumulative exposure
Card operator	1933-35	10.8	32.4
Card operator	1936-1938	6.5	41.9
Clean-up	1939-45	8.8	103.5
Clean-up	1946-48	4.0	115.5

Quantified Personal Measurements

In some instances, quantified personal exposure measurements may be available, e.g. in radiation workers wearing radiation dosimeters (Checkoway et al, 2004). This information is invaluable when it is available, but it is rarely available for historical exposures with the exception of some industries such as the nuclear power industry. Such information can of course be collected prospectively. This is rarely practical for cohort studies of rare diseases with long latency periods (e.g. cancer), but is more appropriate for cohort studies of relatively common conditions. For example, infant cohort studies of respiratory disease frequently prospectively collect information on individual levels of allergen exposure (e.g. Lau et al, 2001).

Quantified personal exposure measurements can also be used in case-control studies to estimate historical exposures. However, a potential problem in this situation is that exposure may have changed over time, or study participants may change their behaviour as a result of having been diagnosed with disease. This has been a particular issue in case-control studies of electromagnetic field exposure and childhood leukemia where it has been argued that current personal exposure measurements may be inferior to "wire code" information (i.e. whether the wiring to the house is underground, or by overhead wires, etc) in estimating historical exposures (Neutra and del Pizzo, 1996).

Example 8.4

Wing et al (1991) conducted a historical cohort mortality study among workers at Oak Ridge National Laboratory, Tennessee. Individual exposures to external penetrating radiation, primarily gamma rays, were measured using pocket ionising chambers from 1943 until June 1944,

film badges from then until 1975, and thermoluminescent dosimeters since 1975. This information was used to estimate individual exposures over time. After accounting for age, birth cohort, socio-economic status, and active worker status, external radiation with a 20-year

exposure lag (i.e. exposures were only considered up until 20 years previously) was associated with an increased risk of death (2.68% increase per 10 mSv cumulative exposure), particularly from cancer (4.94% increase per 10 mSv).

Biomarkers

More recently, there has been increasing emphasis on the use of molecular markers of internal dose (Schulte, 1993). In fact, there are a number of major limitations of currently available biomarkers of exposure (Armstrong et al, 1992), particularly with regard to historical exposures (Pearce et al, 1995). For example, serum levels of micronutrients reflect recent rather than historical dietary intake (Willett, 1990). Some biomarkers are better than others in this respect (particularly markers of exposure to biological agents), but even the best markers of chemical exposures usually reflect only the last few weeks or months of exposure. On the other hand, with some biomarkers it may be possible to estimate historical levels provided that certain assumptions are met. For example, it may be possible to estimate historical levels of exposure to pesticides (or contaminants) from current serum levels provided that the exposure period is known, and the half-life is known. Similarly, information on recent

exposures can be used if it is reasonable to assume that exposure levels (or at least relative exposure levels) have remained stable over time (this may be particularly relevant in occupational studies), and have not been affected by lifestyle changes, or by the occurrence of the disease. However, if the aim is to measure historical exposures, then historical information on exposure surrogates may be more valid than direct measurements of current exposure or dose levels. This situation has long been recognised in occupational epidemiology, where the use of work history records in combination with a job-exposure matrix (based on historical exposure measurements of work areas rather than individuals) is usually considered to be more valid than current exposure measurements (whether based on environmental measurements or biomarkers) if the aim is to estimate historical exposure levels (Checkoway et al, 2004). On the other hand, some biomarkers have potential value in

validation of questionnaires which can then be used to estimate historical exposures. Furthermore, biomarkers of internal dose may have relatively good validity in studies involving an acute effect of exposure.

A more fundamental problem of measuring internal dose with a biomarker is that it is not always clear whether one is measuring the exposure, the biological effect, or some stage of the disease process itself (Saracci, 1984b). Thus the findings may be uninterpretable in terms of the causal association between exposure and disease. When it is known that the biologically effective dose is the most appropriate measure, then the use of appropriate biomarkers clearly has some scientific advantages. However, choosing the appropriate biomarker is a major dilemma, and biomarkers are frequently chosen on the basis of an incomplete or erroneous understanding of the etiologic process (or simply because a particular marker can be measured). An

environmental exposure (e.g. tobacco smoke) may involve hundreds of different chemicals, each of which may produce hundreds of measurable biological responses (there are exceptions to this, of course, such as environmental lead exposure, but most environmental exposure involves complex mixtures). A biomarker typically measures one of the biological responses to one of the chemicals. If the chosen biomarker measures the key etiologic factor, then it may yield relatively good exposure data; however, if a biomarker is chosen which has little relationship to the etiologic component of the complex exposure mixture then the biomarker will yield relatively poor exposure data.

A further major problem with the use of biomarkers is that the resulting expense and complexity may drastically reduce the study size, even in a case-control study, and therefore greatly reduce the statistical power for detecting an association between exposure and disease.

Example 8.5

Ross et al (1992) studied urinary aflatoxin biomarkers and risk of hepatocellular carcinoma as part of an ongoing prospective study of 18,244 middle-aged men in Shanghai. After 35,299 person-years of

follow-up, a nested case-control study was conducted based on the 22 identified cases of liver cancer, and 140 density-matched controls (matched for age and neighbourhood or residence). The cases

of liver cancer were more likely than controls to have detectable concentrations of aflatoxin metabolites (OR = 2.4, 95% CI 1.0-5.9).

Thus, questionnaires and environmental measurements will continue to play a major role in exposure assessment in epidemiology, but biomarkers may be expected to become increasingly useful over time, as new techniques are

developed. The emphasis should be on using "appropriate technology" to obtain the most practical and valid estimate of the etiologically relevant exposure. The appropriate approach (questionnaires, environmental measurements or

biological measurements) will vary from study to study, and from exposure to exposure within the same study, or

within the same complex chemical mixture (e.g. in tobacco smoke).

8.2: Health Status

The type of information required for measuring health status in epidemiological studies may be different from that which is required in clinical practice. As with exposure data, the key issue is that information should be of similar quality for the various groups being compared. For example, suppose that the bladder cancer incidence in a particular geographical area is being compared with national incidence rates; then it would be inappropriate to conduct a pathological review and reclassification of the cases of the cancer identified in the area, since such a reclassification had not been made for the national data and the information would not be comparable. Rather, the cancer cases in the area should be classified exactly as they had been classified in routine national cancer statistics. Thus, the emphasis should be on the comparability of information across the various groups being compared.

The types of health outcome data used in epidemiological studies include: mortality; disease registers; health service records; and morbidity surveys. These can be grouped into data based on routinely collected records, and morbidity data that is collected for a specific epidemiologic study.

Routine Records

Most countries maintain comprehensive death registration systems at the national or regional levels, and cause of

death information for identified deaths can be obtained by requesting copies of death certificates from national, state, or municipal vital statistics offices. In most instances the causes of death are coded by a nosologist trained in the rules specified in the International Classification of Diseases (ICD) volumes compiled by the World Health Organisation. Revisions to the ICD coding are made about every ten years, and in some instances the ICD code for a particular cause of death may change (Checkoway et al, 2004).

Some countries or states also maintain incidence registers for conditions such as cancer, congenital malformations or epilepsy. These have most commonly been established for cancer registration and the International Agency for Research on Cancer (IARC) has been attempting to encourage the establishment of cancer registries and to standardise methods of cancer registration throughout the world (Jensen et al, 1991). Provided that registration is relatively complete, then cancer registrations can provide valuable additional health status information (and increase the number of identified cases) in a cohort study. Furthermore, cancer registries are invaluable for identifying newly diagnosed cases who can be interviewed (while they are still alive) for population-based case-control studies.

Many Western countries have notification systems for occupational diseases. For example, in the United Kingdom the

Surveillance of Work Related and Occupational Respiratory Disease (SWORD) project was established in 1989 as a national surveillance scheme for occupational respiratory disease (Meredith et al, 1991).

As discussed in chapter 9, other routinely collected records can be used

for determining health status in cohort studies, or to create informal "registers" for identifying cases for case-control studies; these include hospital admission records, health insurance claims, health maintenance organisation (HMO) records, and family doctor (general practitioner records).

Example 8.6

Jones et al (1998) performed a record linkage study of pre-natal and early life risk factors for childhood onset diabetes mellitus. They identified 160 boys and 155 girls born during 1965-1986 who had been admitted to hospital in Oxfordshire, England with a diagnosis of diabetes during 1965-1987. For each case, up to eight controls were chosen from records for live births in the same area, matched on sex, year of birth and hospital or place of birth. They

then linked the hospital record for each child to all of that child's hospital records and to his or her mother's maternity record. There were no significant associations between subsequent diabetes and birthweight, gestational age, birthweight for gestational age, maternal age and parity. There were non-significantly increased risks with not breastfeeding (OR=1.33, 95% CI 0.76-2.34) and with diabetes recorded in the mother during pregnancy (OR=5.87,

95% CI 0.90-38.3), and a significantly raised risk with pre-eclampsia or eclampsia during pregnancy (OR=1.48, 95% CI 1.05-2.10). They hypothesized that pre-eclampsia may be the result of an immunogenetic incompatibility between mother and fetus, and that this early immunological disturbance may be related to the incidence of diabetes later in life.

Morbidity Surveys

In some circumstances, routine records may not be available for the health outcome under study, or may not be sufficiently complete or accurate or use in epidemiological studies. Although this could in theory apply to mortality records, more commonly this is an issue for non-fatal conditions, particularly chronic diseases such as respiratory disease and diabetes. Such morbidity surveys may involve clinical examinations (e.g. a clinical history and

peak flow measurements for asthma), more invasive testing (e.g. blood tests for diabetes), questionnaires, or a combination of these methods.

To take the example of asthma, the essential feature of the condition (at least in clinical and epidemiological terms) is variable airflow obstruction which can be reversed by treatment or is self-limiting (Pearce et al, 1998). This poses several problems with the use of

"diagnosed asthma" in asthma prevalence studies, since the diagnosis of "variable airflow obstruction" usually requires several medical consultations over an extended period. It is therefore not surprising that several studies have found the prevalence of physician-diagnosed asthma to be substantially lower than the prevalence of asthma symptoms. Such problems of differences in diagnostic practice could be minimised by using a standardised protocol for asthma diagnosis in prevalence studies. However, this is rarely a realistic option since it requires repeated contacts

between the study participants and physicians, and this is not possible or affordable in large-scale epidemiological studies. Thus, most epidemiological studies must, by necessity, focus on factors which are related to, or symptomatic of, asthma but which can be readily assessed on a particular day. The main options in this regard are symptoms and physiological measurements (Pearce et al, 1998). In particular, standardised symptoms questionnaires have been developed for use in adults (Burney et al, 1994) and children (Asher et al, 1995).

Example 8.7

Dowse et al (1990) studied the prevalence of non-insulin dependent diabetes mellitus (NIDDM) in adults aged 25-74 years in Mauritius. A random sample of 5,892 individuals was chosen and 5,080 (83.4%) participated. They used a 75g oral glucose tolerance test with fasting and 2-h post load blood collection. Glucose tolerance was

classified according to the World Health Organisation (WHO) criteria (World Health Organisation, 1985). The prevalence of NIDDM was similar in men (12.1%) and women (11.7%). Age and sex-standardised prevalence was similar in Hindu Indians (12.4%), Muslim Indians (13.3%), Creoles (10.4%) and Chinese (11.9%). The authors

commented that the findings in Indians were similar to those in other studies of Indian migrant communities, but the findings in Creoles and Chinese were unexpected. "Potent environmental factors shared between ethnic groups in Mauritius may be responsible for the epidemic of glucose intolerance".

Health status can also be measured by more general morbidity and "quality of life" questionnaires. Perhaps the most widely used questionnaire has been the Medical Outcomes Study Short Form (SF-36) (Ware, 1993). This includes scales to measure physical functioning,

role functioning, bodily pain, mental health, and general health perceptions. The SF-36 scales have been widely used in clinical research in a wide variety of populations to assess overall health status.

Summary

Methods of exposure measurement include personal interviews or self-administered questionnaires (completed either by the study participant or by a proxy respondent), diaries, observation, routine records, physical or chemical measurements on the environment, or physical or chemical measurements on the person. Measurements on the person can relate either to exogenous exposure (e.g. airborne dust) or internal dose (e.g. plasma cotinine); the other measurement options (e.g. questionnaires) all relate to exogenous exposures. Traditionally, exposure to most non-biological risk factors (e.g. cigarette smoking) has been measured with questionnaires (either self-administered or interviewer-administered), and this approach has a long history of successful use in epidemiology. Questionnaires may be combined with environmental exposure measurements (e.g. pollen counts, industrial hygiene surveys) to obtain a quantitative estimate of individual exposures. More recently, there has been increasing emphasis on the use of molecular markers of internal dose

(Schulte, 1993). However, questionnaires and environmental measurements have good validity and reproducibility with regard to current exposures and are likely to be superior to biological markers with respect to historical exposures. The emphasis should be on using "appropriate technology" to obtain the most practical and valid estimate of the etiologically relevant exposure.

Similar considerations apply to the collection of information on health status. Once again, it is important that the information obtained should be of comparable quality in the exposed and non-exposed populations. With this proviso, the specific methods used will differ according to the hypothesis and population under study, but the main options include use of routine records (mortality, incidence, hospital admission, health insurance, general practitioner, etc) and the mounting of a special morbidity survey (using clinical examinations, biological testing or questionnaires).

References

- Antó JM, Sunyer J, Newman-Taylor AJ (1996). Comparison of soybean epidemic asthma and occupational asthma. *Thorax* 51: 743-9.
- Armstrong BK, White E, Saracci R (1992). Principles of exposure measurement in epidemiology. New York: Oxford University Press.
- Asher I, Keil U, Anderson HR, et al (1995). International Study of Asthma and Allergies in Childhood (ISAAC): rationale and methods. *Eur Resp J* 8: 483-91.
- Berkman LF, MacIntyre S (1997). The measurement of social class in health studies: old measures and new formulations. Kogevinas M, Pearce N, Susser M, Boffetta P (eds). Socioeconomic factors and cancer. Lyon: IARC, pp 51-64.
- Burney PGJ, Luczynska C, Chinn S, Jarvis D (1994). The European Community Respiratory Health Survey. *Eur Resp J* 7: 954-60.
- Checkoway HA, Pearce N, Kriebel D (2004). Research methods in occupational epidemiology. 2nd ed. New York: Oxford University Press.
- Dement JM, Harris RL, Symons MJ and Shy, CM (1983). Exposures and mortality among chrysotile asbestos workers. Part II: Mortality. *American Journal of Industrial Medicine* 4:421-433.
- Dowse GK, Gareeboo H, Zimmet PZ, et al (1990). High prevalence of NIDDM and impaired glucose tolerance in Indian, Creole, and Chinese Mauritians. *Diabetes* 1990; 39: 390-6.
- Jensen OM, Parkin DM, MacLennan R, et al (1991). Cancer registration: principles and methods. Lyon: IARC.
- Jones ME, Swerdlow AJ, Gill LE, Goldacre MJ (1998). Pre-natal and early life risk factors for childhood onset diabetes mellitus: a record linkage study. *Int J Epidemiol* 1998; 27: 444-9.
- Lau S, Illi S, Sommerfeld C, et al (2001). Early exposure to house-dust mite and cat allergens and development of childhood asthma: a cohort study. Multicentre Allergy Study Group. *Lancet* 2001; 356: 1392-7.
- Liberatos P, Link BG, Kelsey JL (1988). The measurement of social class in epidemiology. *Epidemiologic Reviews* 10: 87-121.
- Meredith SK, Taylor VM, McDonald JC (1991). Occupational respiratory disease in the UK 1989: a report by the SWORD project group. *Br J Ind Med* 1991; 48: 292-8.
- Neutra RR, del Pizzo V (1996). When "wire codes" predict cancer better than spot measurements of magnetic fields. *Epidemiol* 1996; 7: 217-8.
- Pearce N (1988). Multistage modeling of lung cancer mortality in asbestos textile workers. *Int J Epidemiol* 17: 747-52.
- Pearce N (1992). Methodological problems of time-related variables in occupational cohort studies. *Rev Epidem et Santé Publ* 40: S43-S54.
- Pearce N, Beasley R, Burgess C, Crane J (1998). Asthma epidemiology: principles and methods. New York: Oxford University Press.

- Pearce N, Sanjose S, Boffetta P, et al (1995). Limitations of biomarkers of exposure in cancer epidemiology. *Epidemiol* 6: 190-4.
- Polednak AP (1989). Racial and ethnic differences in disease. New York: Oxford University Press.
- Pomare E, Tutengaehe H, Ramsden I, et al (1992). Asthma in Maori people. *NZ Med J* 105: 469-70.
- Raum E, Arabin B, Schlaud M, et al (2001). The impact of maternal education on intrauterine growth: a comparison of former West and East Germany. *Int J Epidemiol* 2001: 30: 81-7.
- Ross RK, Yuan J-M, Yu MC, et al (1992). Urinary aflatoxin biomarkers and risk of hepatocellular carcinoma. *Lancet* 1992; 339: 943-6.
- Saracci R, Simonato L, Acheson ED, et al (1984a). Mortality and incidence of cancer of workers in the man made vitreous fibres producing industry: an international investigation at 13 European plants. *Br J Ind Med* 1984; 41: 425-36.
- Saracci R (1984b). Assessing exposure of individuals in the identification of disease determinants. In: Berlin A, Draper M, Hemminki K, Vainio H (eds). *Monitoring human exposure to carcinogenic and mutagenic agents*. Lyon: IARC.
- Schulte PA (1993). A conceptual and historical framework for molecular epidemiology. In: Schulte P, Perera FP. *Molecular epidemiology: principles and practices*. New York: Academic Press, pp 3-44.
- Vartia M (2001). Consequences of workplace bullying with respect to the well-being of its targets and the observers of bullying. *Scand J Work Environ Health* 27: 63-9.
- Ware JE (1993). *SF-36 Health Survey, Manual and Interpretation Guide*. Boston: The Health Institute.
- World Health Organisation (WHO) (1985). WHO Study Group: Diabetes mellitus. Technical Report Series no 727. Geneva: World Health Organisation.
- Willett W (1990). *Nutritional epidemiology*. New York: Oxford University Press.

CHAPTER 9: Cohort Studies

(In: Pearce N. A Short Introduction to Epidemiology. 2nd ed. Wellington, CPHR, 2005)

As discussed in chapter 2, an incidence study is a subtype of longitudinal study in which the outcome measure is dichotomous (e.g. death or disease incidence). Perhaps the simplest type of incidence study involves "descriptive" analyses using routine mortality or incidence records for a defined geographic population. For example, most countries have comprehensive death registration schemes, as well as regular national censuses, a population register, or other methods of estimating population numbers. These can then be used, as the numerator and denominator respectively, to calculate overall national death rates, as well as the death rates by age-group and gender. In some countries, information may also be available to calculate death rates by other demographic variables such as ethnicity, socio-economic status, employment status, occupation or

geographical area. However, the validity of such analyses may be questionable, because in most countries death certificates (or other routine records such as cancer registration records) are not linked directly to the corresponding population records. Thus, problems may occur if factors such as ethnicity are coded differently on the death records and on the population records. Nevertheless, such "descriptive" analyses, have played a major role in identifying public health problems and suggesting priorities for public health research.

However, the limitations of analyses based on routine records usually mean that a specific "cohort" must be constructed for many epidemiologic studies. In this chapter I discuss the practicalities of conducting a cohort study.

9.1: Defining the source population and risk period

Community-based cohort studies

For studies investigating environmental factors, or general lifestyle (diet, exercise, etc) a cohort study may be based on a particular community which is followed (usually prospectively) over time. For example, a cohort may be based on "all persons aged 20 years or more" living in a particular city or county in a particular year. This would usually require a special survey to be conducted at the start of the follow-up period, with

further surveys being conducted at regular intervals.

More specific cohorts

Cohorts may also be constructed not only on the basis of more specific exposures. Perhaps the most common example of this approach involves studies that are based on workers in a particular factory or industry (Checkoway et al, 2004). Such studies may be based on historical records, enabling follow-up to be conducted retrospectively. Typically, such a

historical cohort study might involve “all workers who worked for at least one month in the factory at any time during 1970-1999”. The list of such workers can be enumerated using personnel records which also provide information on their job titles and departments (which can be used to estimate their historical exposures).

Comparison populations

In community-based cohorts, comparisons are usually made internally between study participants exposed and those not exposed to a particular risk factor (e.g. low dietary beta carotene intake compared with high dietary beta carotene intake).

In studies of specific populations, an internal comparison may still be possible, e.g. by comparing workers with high benzene exposure to those with low benzene exposure. However, in some instances this may not be possible because good individual exposure information is not available (apart from the fact that workers in the factory received high exposure on the average) or because there is not sufficient variation in exposure within the population (e.g. because everyone who worked in the factory had high exposure). In this situation, an external comparison may be made, e.g. with national death rates or cancer registration rates. In this situation, the source population for the study is effectively the national population, and a comparison is being made between the subgroup in the source population that worked in a particular factory (for example) and the entire source population. Ideally the comparison should be made between the exposed group and the source population minus the exposed group (i.e. everyone else in

the country who did not work in the factory). However, this is rarely feasible in practice, and is usually a trivial problem if the exposure is rare. Thus, the comparison is usually made between the exposed group and the national population as a whole.

The risk period

Once the source population has been defined, then the risk period must also be specified. It is important that the risk period is the same for the two or more groups being compared. For example, it would be inappropriate to compare deaths from ischaemic heart disease in two different communities at two different time periods, since there is a continuing decline in IHD mortality, and spurious differences between the communities may be observed if they are not studied over the same risk period.

In a historical cohort study, participants may be followed from some date in the past (e.g. the date the factory opened) up until the present (or some recent date for which death records or cancer registration records are complete). In a prospective cohort study, participants may be followed from the present until some specified future date (e.g. a ten-year follow-up of participants in a recent survey). In both instances, not all study participants will be followed for the entire risk period. For example, someone who moved into the community during the risk period and was “recruited” during a later survey would only be followed from the time of that survey. Similarly, someone who emigrated during the risk period would only be followed until their date of emigration.

Example 9.1

The Renfrew/Paisley study was based on two adjacent urban burghs considered to be typical of the West of Scotland. During 1972-1976, men and women aged between 45 and 64 and identified by door-to-door census as living in Renfrew and Paisley were invited to take part. The response rate was 80% (7,052 men and 8,354 women). Participants completed a questionnaire which

included self-reported smoking history, occupation, address, age, gender, and respiratory symptoms. Study participants were "flagged" at the National Health Service Central Register in Edinburgh and followed for 20 years. Hart et al (2001) reported that high lung cancer mortality risks were seen for manual compared with non-manual workers. The risk reduced when

adjusted for smoking, and reduced further when adjusted for lung function, phlegm and (area) deprivation category. They concluded that the social class difference in lung cancer mortality was explained by poor lung health, deprivation and poor socio-economic conditions throughout life, in addition to smoking.

Example 9.2

Rafnsson et al (2001) studied cancer incidence in a cohort of 1690 flight attendants working with two airline companies in Iceland. The total number of person-years of follow-up was 27,148. Among the 1,532 women flight

attendants, there were 64 cases of cancer, whereas 51.6 were expected on the basis of national cancer incidence rates (RR=1.2). There was a particularly elevated risk for breast cancer in those who had been

hired in 1971 or later and therefore had had the heaviest exposure to cosmic radiation at a young age (RR=4.1). The authors concluded that the association may be due to cosmic radiation or disturbance of circadian rhythm.

9.2: Measuring exposure

As discussed in chapter 8, there are a variety of possible methods for measuring exposure in cohort studies. These include routine records, questionnaires, environmental measurements, Job-Exposure-Matrices (JEM), quantified personal measurements, and biomarkers of exposure.

Ideally, exposures should be measured continuously, or at least at regular intervals, through the risk period (i.e. the period of follow-up). For some risk factors (e.g. for demographic factors such as age, gender and ethnicity), the risk factor status is unlikely to change during the risk period, and can simply be ascertained at baseline. For other exposures that do change over time (e.g. smoking, diet, occupational

exposures) regular surveys, or regular examination of routine records, may be desirable to update the exposure information. However, in many studies this is not feasible and information is only collected in a baseline survey; it is then necessary to assume that the exposure level (e.g. serum cholesterol level) has not changed meaningfully during the subsequent follow-up.

In occupational studies, more detailed exposure information may be available through the combination of personnel records (which include changes of job title and department) and Job-Exposure-Matrices (JEMs) based on workplace exposure surveys and/or personal measurements in a subgroup of the workforce (see chapter 8).

Example 9.3

Prescott et al (2004) studied vital exhaustion (fatigue, hopelessness and depression) as a risk factor for ischaemic heart disease (IHD) in 4084 men and 5479 women in Copenhagen. The study was based on participants in the Copenhagen City Heart Study, and the analyses were based on 10,135 people who attended the third follow-up examination in 1991-1993. Cardiovascular risk factors were assessed by a self-administered

questionnaire checked with the participant by trained staff, and by various laboratory tests. Vital exhaustion was assessed using a 17-item questionnaire. Participants were followed until 31 December 1997 for fatal and non-fatal IHD, with the information being obtained from the National Board of Health and National Hospital Discharge Register respectively. Subjects with self-reported and verified IHD prior to enrolment were excluded. During

follow-up, 483 experienced an IHD event, of which 25% were fatal, and 1559 subjects died from all causes. All but 4 of the 17 items were significantly associated with IHD with significant relative risks ranging from 1.36 to 2.10. The RR for IHD in those with a vital exhaustion score of 10 or more was 2.57 (95% CI 1.65-4.00) and this altered little after adjustment for biological, behavioural and socioeconomic risk factors.

9.3 Follow-up

Vital status ascertainment

In some instances, particularly in community based studies, follow-up may involve regular contact with the study participants, including repeated surveys of health status. Perhaps more commonly, follow-up may not involve further contact with the study participants, but may be done by routine record linkage.

For example, study participants may be followed over time by linking the study information with national death records, or incidence records (e.g. a national cancer registry) as well as with other record systems (e.g. social security records, drivers license records) to confirm vital status in those who are not found to have died during the follow-up period.

Although most developed countries have complete systems of death registration, and it is easy in theory to identify all deaths in a particular cohort, this may not be so straightforward in practice. For example, many countries do not have national identification numbers and record linkage may have to be done on the basis of name and date of birth. This may not be infallible because of differences in spelling of names, or inaccuracies in date of birth, but various record linkage programmes are available to identify "near matches" (Jones and Sujansky, 2004). These will be ineffective, however, for people who have changed their name, e.g. because of marriage.

A further problem is that some countries do not have national death registrations, and these may be done on a regional or state basis instead, making it necessary to search multiple registers. Since 1979 a National Death Index for the United States has been compiled and computerized and is available for vital status tracing (Wentworth et al, 1983).

Just because someone has been not been identified in death records, this does not mean that they are still alive and "at risk" since they may have emigrated or may not have been identified in death registrations for some other reason. It is therefore desirable to confirm that they are alive using other record sources such as drivers license records, voter registrations, social security records, etc. In the United States, the Social Security Administration (SSA) records have been frequently used in the past, and in Great Britain the Central Record Office of the Ministry of Pensions and National Insurance is the analogous tracing source (Checkoway et al, 2004).

Coding of the disease outcome

It is not only necessary to determine if and when an event such as a death or hospital admission occurred. It is also necessary to verify, for example, the cause of death, or the cause of a hospital admission. Coding of causes of death should be performed by a nosologist trained in the rules specified by the International Classification of Diseases (ICD) volumes compiled by the World Health Organisation. In many countries this is done routinely for national death

registration records, and it is not necessary (or desirable) to recode death registrations for a specific study. However, the ICD codes have changed over time, and when using routine death registration records it is necessary to be aware of which ICD revision was in effect at the time of death.

Person-time

In a study of a specific population, e.g. workers in a particular factory, participants may enter the study on the date that the study starts (1/1/70), or the date that they first

meet the eligibility criteria (i.e. employment for one month), whichever is the latest date. If they started working in the factory after the start of the study, then they would only start being followed on the date they started work (or a subsequent date when they met the eligibility criteria).

They stop contributing person-time when they die (or are diagnosed with the disease in an incidence study), emigrate, they are lost to follow-up, or the study finishes (31/12/99) whichever is the earliest.

Example 9.4

Munk Nielsen et al (2003) studied long-term mortality after poliomyelitis by identifying a group of 5,977 patients diagnosed with poliomyelitis in Copenhagen between 1919 and 1954. This involved a review of more than 80,000 consecutive hospital records for Blegdamshospitalet which served as the primary centre for diagnosing and treating patients with acute poliomyelitis in the area of greater Copenhagen. Information extracted from the records included name, sex, date and place of birth, date of admission and discharge, and details of the acute severity of the case.

Since 1 April 1968, all Danish citizens have been given a unique identification number, which is recorded in the Danish Civil Registration System (CRS). The cohort was linked to the CRS to identify individual CRS numbers which were then used to identify deaths in the Danish Cause-of-Death Register. Patients not identified in the CRS were believed to have died or emigrated before 1 April 1968 and for these patients the Cause-of-Death Register was searched for their name and date of birth.

Patients were followed from the initiation of the Cause-of-Death Register in 1943 or the month after the hospital

discharge (whichever came later) until the date of death, emigration or 1 May 1997 (whichever came earlier).

There were 1295 deaths compared with an expected number of 1141 (SMR 1.14, 95% CI 1.07-1.20). Excess mortality was restricted to polio patients with a history of severe paralysis of the extremities (SMR = 1.69, 95% CI 1.32-2.15) or patients who had been treated for respiratory failure during the epidemics (SMR = 2.71; 95% CI 2.18-3.37).

Summary

Cohort studies provide the most comprehensive approach for evaluating patterns of exposure and disease, since they involve studying the entire source population (assuming that there is a 100% response rate) over the entire risk period.

Thus, the cohort design ideally includes all of the relevant person-time experience of the source population over the risk period. A cohort study may be based on a particular community (e.g. a geographical community), or on a more specific population defined by a particular exposure (e.g. workers in a particular factory). In both instances, an internal comparison would ideally be made between those participants

exposed and those participants not exposed to a particular risk factor. However, in some instances, all of the study participants may be exposed, or valid individual exposure information may not be available, and it may be necessary to make an external comparison, e.g. with national mortality rates (in which case the national population comprises the source population for the study). It is important that any comparisons are made over the same risk period, and that follow-up is as complete as possible. The basic effect measures in a cohort study are the rate ratio and risk ratio. Methods of data analysis for these effect measures are described in chapter 12.

References

- Checkoway HA, Pearce N, Kriebel D (2004). *Research methods in occupational epidemiology*. 2nd ed. New York: Oxford University Press.
- Hart CL, Hole DJ, Gillis CR, et al (2001). Social class differences in lung cancer mortality: risk factor explanations using two Scottish cohort studies. *Int J Epidemiol* 30: 268-74.
- Jones L, Sujansky W (2004). *Patient data matching software: a buyers guide for the health conscious*. Okaland, CA: California HealthCare Foundation (www.chcf.org).
- Munk Nielsen N, Rostgaard K, Juel K, et al (2003). Long-term mortality after poliomyelitis. *Epidemiol* 14: 355-60.
- Prescott E, Holst C, Gronbaek M, et al (2004). Vital exhaustion as a risk factor for ischaemic heart disease and all-cause mortality in a community sample: a prospective study of 4084 men and 5479 women in the Copenhagen City Heart Study. *Int J Epidemiol* 33: 990-7.
- Rafnsson V, Tulinius H, Jónasson JG, Hrafnkelsson J (2001). Risk of breast cancer in female flight attendants: a

population-based study (Iceland).
Cancer Causes and Control 12: 95-
101.

Wentworth DN, Neaton JD, Rasmussen
WL (1983). An evaluation of the

Social Security Administration Master
Beneficiary Record file and the
National Death Index in the
ascertainment of vital status. Am J
Public Health 73; 127-1274.

CHAPTER 10: Case-control Studies

(In: Pearce N. A Short Introduction to Epidemiology. 2nd ed. Wellington, CPHR, 2005)

As discussed in chapter 2, the only conceptual difference between a full cohort study based on a specified source population and risk period, and an (incidence) case-control study based on the same source population and risk period, is that the latter involves outcome-specific samples of the source population, rather than an analysis of the entire source population. There is

usually little loss of precision compared to a full cohort study, and there may be considerable savings in terms of time and expense, particularly if the study disease is rare or has a long induction time.

In this chapter I discuss the practicalities of conducting an (incidence) case-control study.

10.1: Defining the source population and risk period

An incidence case-control study should be based on a specified source population and risk period. The task in such a population-based case-control study is then to identify all cases of the outcome under study that are generated by the source population over the risk period. Controls are then sampled at random from the source population.

In some instances, cases may be identified from a particular disease register which is not comprehensive with respect to the population of any

defined geographical area. This may be a formal register (such as a Cancer Register) or a similar data source (e.g. admission records for a particular hospital). In such a registry-based study the task is to identify the source population for the register (e.g. all persons who would have been admitted to the hospital if they had developed the disease under study). This obviously poses more problems in the appropriate selection of controls than is the case for a population-based study; these issues are discussed in more depth below.

Example 10.1

Bigert et al (2001) studied myocardial infarction (MI) among professional drivers. The source population comprised all men aged 45-70 years free of previous MI and living in Stockholm County during 1992-1993. Cases of first MI generated by this source population and risk period were identified from three sources: the medical care units at the 10 emergency hospitals within the Stockholm County (87% of the cases), other hospital

units (1% of the cases, obtained from a computerized hospital discharge register), or death certificates from the Causes of death Register at Statistics Sweden (12% of the cases). Controls were selected at random from a computerized population register, stratified for sex, 5-year age-group, hospital catchment area and year of enrolment in the study (1992 or 1993). The 1,067 cases and 1,482 controls completed a postal

questionnaire (for fatal cases the questionnaires were completed by next-of-kin). Of the cases, 45 (4.2%) had worked as a bus driver, compared with 31 (2.1%) of the controls, yielding an odds ratio of 2.14 (95% CI 1.34-3.41). The corresponding odds ratios for taxi drivers and truck drivers were 1.88 (95% CI 1.19-2.98) and 1.66 (95% CI 1.22-1.26) respectively. Adjustment for potential confounders gave lower odds ratios: 1.49, 1.34 and 1.10 respectively.

10.2: Selection of cases

In a population-based study the first step in the selection of cases is to attempt to ascertain all cases generated by the source population over the risk period (Checkoway et al, 2004). If complete case ascertainment is not achieved, then the relative risk estimate (odds ratio) will not necessarily be biased unless case ascertainment is associated with exposure history, e.g. if people who are prescribed a particular drug receive more intensive medical

screening and are therefore more likely to be diagnosed with a non-fatal myocardial infarction.

In a registry-based study the case-group usually consists of all incident cases occurring in the registry during the risk period. The "registry" could consist of a formal population-based registry (e.g. a cancer registry or birth defects registry), or could involve an ad hoc "registry", e.g. based on admission records for the major hospitals in a city.

Example 10.2

Mian et al (2001) studied homicide in Orangi, the largest squatter settlement in Karachi with an estimated population of 1.2 million. They defined the cases as individuals who lived in Orangi and were killed in Orangi between January 1994 and January 1997, due to intentional violence, by firearms, sharp or blunt trauma. Cases were identified in the 15 neighbourhoods (out of 103 in total in Orangi) which field workers identified as the highest violence neighbourhoods. Field workers identified households where they knew someone had been killed; in a few neighbourhoods they also contacted other social organisations in the community to identify further cases. Controls were selected from a random sample of households enrolled in a related study conducted at the same time in the same 15 neighbourhoods. For

both cases and controls, the interviews were conducted with their wife, or if she was inaccessible or unwilling it was conducted with the wife of the head of the household. People who were killed were 34 times more likely to have attended all political processions (29% versus 1%, odds ratio (OR) = 34, 95% CI 4-749), 19 times more likely to have attended political meetings (31% versus 2%, OR = 19, 95% CI 4-136), and 17 times more likely to have held an important position in a political party (29% versus 2%, OR = 17, 95% CI 3-120). The authors concluded that homicide in Orangi was political and that efforts to build trust between ethnic groups and to build legitimacy for non-violent forms of conflict resolution are important steps to limit future violence.

10.3: Selection of controls

Control sampling options

As discussed in chapter 2, there are three main options for selection of controls: (i) cumulative sampling involves selecting controls from those who do not experience the outcome during the risk period (i.e. the survivors) and will estimate the incidence odds; (ii) case-cohort sampling involves selecting controls from the entire source population and will estimate the risk ratio; (iii) density sampling involves selecting controls longitudinally throughout the course of the study and will estimate the rate ratio. Density sampling is therefore usually the preferred approach since

the rate ratio is usually the effect measure of interest. Fortunately, although case-control studies have traditionally been presented in terms of cumulative sampling (e.g. Cornfield, 1951), most case-control studies actually involve density sampling (Miettinen, 1976), often with matching on a time variable (such as calendar time and/or age), and therefore estimate the rate ratio without the need for any rare disease assumption (Pearce, 1993). In particular, the "standard" population-based case-control design in which all cases occurring in a country (or state or city) in a particular year are compared with a control sample of all other people

living in the same country during the same year, actually involves density sampling with calendar year as the "time" matching variable (possibly with additional matching on the additional "time" variable of age).

Sources of controls

In a population-based case-control study, controls are usually sampled at random from the entire source population (perhaps with matching on factors such as age and gender). In some instances, it may be necessary to restrict the source population in order to achieve valid control sampling. For example, if controls are to be selected from voter registration rolls, and these are known to be less than 100% complete for the geographical area under study, then the source population might be restricted to persons appearing on the voter registration roll, and cases that were not registered to vote would be excluded; controls would then be sampled from this redefined source population by taking a random sample of the roll.

In registry-based studies, selection of controls may not be so straightforward because the source population may not be so easy to define and enumerate. For example, if there are two major hospitals in a city, and a study is based on lung cancer admissions in one of them during a defined risk period, then the source population is "all those who would have come to this hospital for treatment if they had developed lung cancer during this risk period". This population may be difficult to define and enumerate, particularly if cases may also be referred from smaller regional hospitals. The best solution is usually to define a more specific source population (e.g. all people living in the city) and to attempt to identify all cases generated by that source population, e.g. by including

admissions from all major hospitals in the city and excluding cases who do not live in the city; controls can then be sampled from that defined source population.

If it proves impossible to define and enumerate the source population, then one possibility is to select controls from people appearing in the same "register" for other health conditions (e.g. admissions to the hospital for other causes). This may not only produce a valid sample of the "source population", but may also have advantages in making the case and control recall more comparable (Smith et al, 1988). However, it may result in bias if the other health conditions are also caused (or prevented) by the exposure under study (Pearce and Checkoway, 1988). For this reason, the population-based approach is preferable, although registry-based studies may still be valuable when population-based studies are not practicable, provided that careful consideration is given to possible sources of bias.

Matching

In some instances it may be appropriate to match cases and controls on potential confounders (e.g. age and gender). This can be done by 1:1 matching (e.g. for each case, choose a control of the same age and gender) or by frequency matching (e.g. if there are 25 male cases in the 30-34 age-group then choose the same number of male controls for this age-group). It is important to emphasize, however, that this will not remove confounding in a case-control study, but will merely facilitate its control in the analysis. For example, in a case-control study of lung cancer, the cases will generally be relatively old whereas a random general population control sample will be relatively young. This may lead to inefficiencies when age

is controlled in the analysis since the older age-groups will contain many cases and few controls, whereas the younger age-groups will contain many controls and few cases. Matching on age will ensure that there are approximately equal numbers of cases and controls in each age-strata and will thereby improve the precision of the effect estimates (given a fixed number of cases and controls). However, it will not remove confounding by age – it merely makes it easier to control in the analysis (Checkoway et al, 2004).

It is also important to emphasize that if “pair” matching (i.e. 1:1 matching) has been done, then it is important to control for the matching factors in the analysis, but that this need not involve a “matched analysis”. For example, if pair matching has been done on age and gender, then it is important to control for age and gender in the analysis, but this

can be done with simple stratification on age (e.g. by five-year age-groups) and gender and it is not necessary to retain the 1:1 matched pairs in the analysis (Rothman and Greenland, 1998).

There are also potential disadvantages of matching. In particular, matching may actually reduce precision in a case-control study if it is done on a factor that is associated with exposure but is not a risk factor for the disease under study and hence is not a true confounder (Rothman and Greenland, 1998). Furthermore, matching is often expensive and/or time consuming. For these reasons, it is usually sufficient, and preferable, to only match on basic demographic factors such as age and gender, and to then control for other potential confounders (along with age and gender) in the analysis (Checkoway et al, 2004).

Example 10.3

Cole et al (2000) studied time urgency and risk of non-fatal myocardial infarction (MI) in a study of 340 cases and an equal number of age, sex and community-matched controls. Cases were identified from admissions to the coronary or intensive care units of six suburban Boston hospitals between 1 January 1982 and 31 December 1983. Those eligible for inclusion were white men and

women under 76 years old living in the Boston area with no previous history of MI. For each case, a control subject of the same sex and age (± 5 years) was selected at random from the residents' list of the town in which the patient resided. Each subject was interviewed in his or her home by one of two trained nurse interviewers approximately 8 weeks after discharge from the hospital. A sense of

time urgency/impatience was ascertained using four items from the 10-item Framingham Type A scale. A dose-response relation was apparent among subjects who rated themselves higher on the four-item urgency/impatience scale with a matched odds ratio for non-fatal MI of 4.45 (95% CI 2.20-8.99) comparing those with the highest rating to those with the lowest.

10.4: Measuring exposure

Once the cases and controls have been selected, information on previous exposures is then obtained for both groups. As discussed in chapter 8, there are a variety of possible methods for measuring exposure in case-control studies. In some instances this may be from historical records, e.g. personnel records that contain work history information.

Perhaps more commonly, exposure information may be obtained from questionnaires. It is this latter feature of case-control studies which has left them

open to criticism as being particularly prone to bias, e.g. because the recall of past exposures (e.g. eating meat, drinking alcohol, spraying pesticides) may be different between cases of disease and healthy controls. However, collecting exposure information from questionnaires is not an inherent feature of case-control studies, and is sometimes also a feature of cohort studies. Thus, there is nothing inherently biased in the case-control design; rather what is important is the validity of the exposure information that is collected, whatever study design is employed.

Summary

The only conceptual difference between a full cohort study based on a specified source population and risk period, and an (incidence) case-control study based on the same source population and risk period, is that the latter involves outcome-specific samples of the source population, rather than an analysis of the entire source population. There is usually little loss of precision compared to a full cohort study, and there may be considerable savings in terms of time and expense, particularly if the study disease is rare or has a long induction time.

The key feature of good case-control study design is that the study should be based on a specified source population

and risk period. The tasks are then to: (i) identify all cases generated by the source population over the risk period; (ii) select a random sample of controls from the source population over the risk period (ideally by density matching); (iii) obtain exposure information from cases and controls in a standardised and unbiased manner.

The standard effect estimate in a case-control study is the odds ratio. If controls are selected by density matching, then the odds ratio will estimate the incidence rate ratio (in the source population and risk period) in an unbiased manner without the need for any rare disease assumption. Methods of data analysis for odds ratios are described in chapter 12.

References

- Bigert C, Gustavsson P, Hallqvist J, et al (2003). Myocardial infarction among professional drivers. *Epidemiol* 14: 333-9.
- Checkoway HA, Pearce N, Kriebel D (2004). *Research methods in occupational epidemiology*. New York: Oxford University Press.
- Cornfield J (1951). A method of estimating comparative rates from clinical data: applications to cancer of the lung, breast and cervix. *JNCI* 11: 1269-75.
- Mian A, Mahmood SF, Chotani H, Luby S (2001). Vulnerability to homicide in Karachi: political activity as a risk factor. *Int J Epidemiol* 31: 581-5.
- Miettinen OS (1976). Estimability and estimation in case-referent studies. *Am J Epidemiol* 103: 226-35.
- Pearce N (1993). What does the odds ratio estimate in a case-control study? *Int J Epidemiol* 22: 1189-92.
- Pearce N, Checkoway H (1988). Case-control studies using other diseases as controls: problems of excluding exposure-related diseases. *Am J Epidemiol* 127: 851-6.
- Rothman KJ, Greenland S (1998). *Modern epidemiology*. 2nd ed. Philadelphia: Lippincott-Raven.
- Smith AH, Pearce N, Callas PW (1988). Cancer case-control studies with other cancers as controls. *Int J Epidemiol* 17: 298-306.

CHAPTER 11: Prevalence Studies

(In: Pearce N. A Short Introduction to Epidemiology. 2nd ed. Wellington, CPHR, 2005)

As discussed in chapter 3, incidence studies are usually the preferred approach, but may be time consuming and expensive, and it may be difficult to identify incidence cases of non-fatal chronic conditions such as diabetes. In particular, some degenerative diseases (e.g. chronic bronchitis) may have no clear point of onset.

Thus, in some settings (e.g. developing countries) and for some conditions (e.g. chronic non-fatal disease) prevalence studies may be the only realistic option. Furthermore, in some instances we may be more interested in factors that affect the current burden of disease in the population (i.e. prevalence) rather than disease incidence.

Examples of prevalence surveys include general households surveys conducted by government agencies (e.g. Ministry of Health, 1999), more focussed general population surveys (e.g. Australian Bureau of Statistics, 1998) such as the National Health and Nutrition Examination Survey (NHANES) (Kuczmarski et al, 1994), international surveys of the prevalence of conditions such as asthma (Burney et al, 1994; Asher et al, 1995), and surveys in populations with specific exposures (e.g. surveys of asthma in children living on farms (Braun-Fahrlander et al, 1999)).

In this chapter I discuss the practicalities of conducting a prevalence study.

11.1: Defining the source population

Prevalence studies usually involve surveys in a source population defined by a geographic region or a particular exposure (e.g. an industry or factory). As with an incidence study, it is important that this source population is well-defined and that a high response rate is obtained.

In a prevalence study, disease prevalence is measured at a specific point in time, rather than over a specified risk period. However, this

“point in time” may not necessarily be the same “date” for each study participant. For example, studies of congenital malformations usually involve measuring the prevalence of congenital malformations at birth. Thus the source population may be “all babies born in this city during 2004” and the “time” at which prevalence is measured may be “birth” which will be a different date and time for each member of the source population.

Example 11.1

Wilks et al (1999) conducted a survey of the prevalence of diabetes in the population of Spanish Town, Jamaica. A random population sample was recruited by door-to-door canvassing (n=1,303) and oral

glucose tolerance testing was conducted after an overnight fast (response rate = 60%). The prevalence of Type 2 diabetes mellitus was 15.7% among women and 9.8% among men. The sex patterns were consistent with the

fourfold excess of diabetes in women compared to men, but obesity could not entirely account for the high prevalences observed which exceed those previously reported among European populations.

11.2: Measuring health status

Prevalence studies differ from incidence studies in that the measurement of health status most commonly involves a morbidity survey, rather than identifying incident cases through routine records (e.g. hospital admissions or cancer registration records). Methods that can be used for such surveys have already been discussed in chapter 8 and will only be considered briefly here.

As discussed in chapter 8, methods of measuring disease status, that are most appropriate in clinical practice may not be appropriate or applicable in epidemiologic surveys. Furthermore, the criteria for deciding the most valid method to use may differ between clinical practice and epidemiological surveys. In the clinical setting the emphasis is often on the positive predictive value of a test, which depends in turn on the sensitivity, specificity, and the underlying population prevalence of the disease.

In fact, if we are using a particular method to measure the prevalence of a disease, and:

S_n = sensitivity

S_p = specificity

P is the true prevalence of the disease in the source population

then the observed prevalence that will be obtained in the survey is:

$$S_n P + (1 - S_p) (1 - P) \\ = P (S_n + S_p - 1) + (1 - S_p)$$

therefore if two populations are being compared, and their true prevalences (according to the gold standard) are P_1 and P_0 respectively, then the observed difference in prevalence between the two centres is:

$$(P_1 - P_0)(S_n + S_p - 1)$$

The expression $(S_n + S_p - 1)$ is Youden's Index. When this is equal to 1 (which only occurs when the sensitivity and specificity are both 1) then the observed difference in

prevalence will be exactly equal to the true difference in prevalence. More commonly, Youden's Index will be less than 1 and the observed prevalence difference will be reduced accordingly, e.g. if Youden's Index is 0.75 then the observed prevalence difference will be 0.75 times the true prevalence difference. Youden's Index therefore provides the most appropriate

measure of the validity of a particular question or technique in prevalence comparisons (Pekkanen and Pearce, 1999).

In this respect, basic symptom questionnaires may often perform better than supposedly more "objective" measures such as bronchial responsiveness testing (Pearce et al, 1998).

Example 11.2

Table 11.1 shows hypothetical data from a study of asthma prevalence in childhood. The true prevalence rates were 40% in the exposed group, and 20% in the non-exposed group; the true prevalence difference was thus 20%. If 20% of asthmatics are incorrectly classified as non-asthmatics (i.e. a sensitivity of 0.80), and 10% of non-asthmatics are incorrectly classified as asthmatics (i.e. a specificity of 0.90), then the observed prevalences will be 38%

and 24% respectively (table 11.1); the observed prevalence difference will then be 14% (instead of the true value of 20%). The net effect is to bias the prevalence difference towards the null value of zero. The extent of the bias is related to Youden's index: this is $0.80+0.90-1.0=0.7$, and the observed prevalence difference of 14% is 0.7 times the true value of 20%. If the sensitivity and specificity had been perfect (1.0) then Youden's Index would have been 1.0 and there

would have been no diminishment in the observed prevalence difference; on the other hand, if the sensitivity and specificity had been no better than chance (e.g. both equal to 0.5) then Youden's Index would have been zero, and the expected value of the observed prevalence difference also have been zero (although the observed value might be different from zero due to chance variation).

Table 11.1

Hypothetical data from a prevalence study in which 20% of asthmatics and 10% of non-asthmatics are incorrectly classified

	Actual		Observed	
	Exposed	Non-exposed	Exposed	Non-exposed
Asthmatics	40	20	32 + 6 = 38	16 + 8 = 24
Non-asthmatics	60	80	54 + 8 = 62	72 + 4 = 76
Total	100	100	100	100
Prevalence	40%	20%	38%	24%

11.3: Measuring exposure

As discussed in chapter 8, there are a variety of possible methods for measuring exposure in prevalence studies. These include questionnaires, biological measurements, and examination of historical records (e.g. personnel and work history records).

In a full prevalence study, exposure is measured in all members of the source population. In a prevalence case-control study, exposure information is obtained for the cases and for a control sample of non-cases (chapter 3). Thus, a prevalence case-control study can be based on routine records (see example 3.2) or as a second phase of a specific prevalence survey.

Whereas incidence case-control studies involve at least three possible methods of selecting controls, in a prevalence case-control study there is only one valid option, i.e. controls should be selected at random from the non-cases. For both groups, information on historical and current exposures may be obtained, as well as information on potential confounders.

It is important to emphasize that although a prevalence study involves measuring disease status at one point in time, information can be collected on historical exposures. For example, a prevalence survey of bronchitis might involve assessing whether a person has "current bronchitis" on a particular day, but exposure information could be

collected for both current smoking as well as smoking history.

Example 11.3

Guha Mazumder et al (2000) studied arsenic in drinking water and the prevalence of respiratory effects in West Bengal, India. A cross-sectional survey involving 7,683 participants of all ages was conducted in an arsenic-affected region between April 1995 and March 1996. The source population was based on two areas of the arsenic-affected districts south of Calcutta. A convenience sampling strategy was used in which the field team went to the centre of each village and selected the most convenient hamlet to begin sampling; all

household members were invited to participate and sampling continued from house to house until sufficient numbers had been recruited. Participants were clinically examined and interviewed, and the arsenic content of their current primary drinking water source was measured. There were few smokers and analyses were confined to non-smokers (6,864 participants). Among both males and females, the prevalence of cough, shortness of breath, and chest sounds (crepitations and/or rhonchi) in the lungs rose with increasing

arsenic concentrations in drinking water. In participants with arsenic-related skin lesions, the age-adjusted prevalence odds ratios for cough were 7.8 for females (95% CI 3.1-19.5), and 5.0 for males (95% CI 2.6-9.9); the corresponding findings for chest sounds were 9.6 (95% CI 4.0-22.9) and 6.9 (95% CI 5.8-92.8), and those for shortness of breath were 23.3 (95% CI 5.8-92.8) and 3.7 (95% CI 1.3-10.6). The authors concluded that these results add to evidence that long-term ingestion of arsenic can cause respiratory effects.

Summary

Incidence studies are usually the preferred approach, but in some settings and for some conditions prevalence studies are the only option. Furthermore, in some instances we may be more interested in factors that affect the current burden of disease in the

population rather than disease incidence. The conduct of a prevalence study is (at least in theory) relatively straightforward. A source population is defined, and at one point in time the prevalence of disease is measured in the population. Exposure information is then

obtained for all members of the source population (a prevalence study), or for all cases of the disease under study and a control sample of the non-cases (a prevalence case-control study). The

standard effect estimate in a prevalence study is the odds ratio. Methods of data analysis for odds ratios are described in chapter 12.

References

- Asher I, Keil U, Anderson HR, et al (1995). International study of asthma and allergies in childhood (ISAAC): rationale and methods. *Eur Resp J* 8: 483-91.
- Australian Bureau of Statistics (1998). National Nutrition Survey: Nutrient Intakes and Physical Measurements, Australia, 1995. Canberra: Australian Bureau of Statistics.
- Braun-Fahrlander CH, Gassner M, Grize L, et al (1999). Prevalence of hay fever and allergic sensitization in farmer's children and their peers living in the same rural community. *Clin Exp Allergy* 29: 28-34.
- Burney PGJ, Luczynska C, Chinn S, Jarvis D (1994). The European Community Respiratory Health Survey. *Eur Resp J* 7: 954-60.
- Checkoway HA, Pearce N, Kriebel D (2004). *Research methods in occupational epidemiology*. 2nd ed. New York: Oxford University Press.
- Guha Mazumder DN, Haque R, Ghosh N, et al. Arsenic in drinking water and the prevalence of respiratory effects in West Bengal, India (2000). *Int J Epidemiol* 29: 1047-52.
- Kuczmarski RJ, Flegal KM, Campbell SM, Johnson CL (1994). Increasing Prevalence of Overweight Among US Adults. The National Health and Nutrition Examination Surveys, 1960 to 1991. *JAMA* 1994; 272: 205-211
- Ministry of Health (1999). Taking the pulse: the 1996/1997 New Zealand Health Survey. Wellington: Ministry of Health.
- Pearce N, Beasley R, Burgess C, Crane J (1998). *Asthma epidemiology: principles and methods*. New York: Oxford University Press.
- Pekkanen J, Pearce N. Defining asthma in epidemiological studies. *Eur Respir J* 1999; 14: 951-7.
- Wilks R, Rotimi C, Bennett F, et al (1999). Diabetes in the Caribbean: results of a population survey from Spanish Town, Jamaica. *Diabetic Medicine* 1999; 16: 875-83.

Part IV

Analysis and Interpretation of Studies

CHAPTER 12: Data Analysis

(In: Pearce N. A Short Introduction to Epidemiology. 2nd ed. Wellington, CPHR, 2005)

In this chapter I describe the basic principles of data analysis in epidemiologic studies including the estimation of effects and calculation of confidence intervals while controlling for potential confounders. I only cover the basic methods for dichotomous exposures and dichotomous health

outcomes (chapters 2 and 3) and I do not consider more complex study designs (chapter 4). Readers requiring a more formal and detailed statistical presentation are referred to standard texts (particularly Rothman and Greenland, 1998).

12.1: Basic Principles

Data Management

With the rapid advances in computer technology in recent years, almost any epidemiological study can be analysed on a personal computer (PC). In addition, a wide variety of software is available for data entry, data analysis and graphical presentation of data on PCs (much of which is not available for mainframe computers). One particularly useful package is EPI-INFO (Dean et al, 1990), which is available through WHO (Geneva) and CDC (Atlanta) and can be downloaded from <http://www.cdc.gov/epiinfo>. This package is particularly useful for data entry and editing, and can be used on small laptop computers in the field as well as on desktop computers. However, the same facilities are available in many other packages, some of which are more sophisticated both statistically and in terms of data management (e.g. Stata (Hills and De Stavola, 2002)). A catalogue of

epidemiological resources, including epidemiological software which is available free of charge, or at minimal cost, has been produced by the Epidemiology Monitor, and this publication also has a regular feature reviewing such software (see the Epidemiology Monitor Website at <http://www.epimonitor.net/>). There is also an excellent epidemiology Excel spreadsheet (Episheet) available, which can be used to do most of the analyses described in this chapter (Rothman, 2002). It can be downloaded from <http://www.oup-usa.org/epi/rothman/>.

Given the huge amount of work usually involved in collecting data for epidemiologic studies, it is essential to examine the raw data very carefully for errors and to make every attempt to avoid errors in the transfer of data from questionnaires onto the computer. In most cases, the first step is to translate some of the information into numerical (or alphabetical) codes, following a set

of coding instructions that should have been prepared prior to data collection. For instance, a detailed occupational history may have been taken in a semi-narrative form, and must be subsequently coded. It is usually preferable to do this when entering the data directly onto a PC, since this minimizes transcription errors.

Once the data are coded and entered, programmes should be run that seek strange data, contradictions, and impossible data (e.g. a systolic blood pressure of 40 mm Hg). These programmes should not be restricted to a search for logic errors or impermissible symbols. They should include also procedures that identify values that lie outside plausible limits. The values being queried should be listed, and decisions on how the "errors" are dealt with should be documented. With many packages, this process can be conducted during the actual data entry since the range of permissible values (for numeric variables) or legal codes (for alphanumeric variables) can be specified, as well as variables which must not be left blank, conditional jumps (e.g. if the answer is "NO" the computer skips to the next relevant question), repeat fields (so that the value of a variable is set by default to that of the last record entered or displayed), and logical links between variables. The best method of data checking is to enter all of the data twice, and to compare the two files for discrepancies. This approach, combined with extensive edit checks at the time of data entry, should minimize errors.

Even with double data entry and sophisticated checking procedures, errors may occur, and it is therefore important to run further edit checks before data analysis begins. It is particularly important to finish all edit checks and to have a final version of the data file before any data analysis is

done, both to avoid confusion, and also to avoid any possibility of the data coding and checking being influenced by the results of preliminary analyses. Once the data have been entered and edited, there is usually a major task of data management. This typically involves the use of a computer package to transform the data, compute new variables, and prepare new files suitable for statistical analysis.

Data Analysis

The basic aim of the analysis of a single study is to estimate the effect of exposure on the outcome under study while controlling for confounding and minimizing other possible sources of bias. In addition, when confounding and other sources of bias cannot be removed, then it is important to assess their likely strength and direction. This latter task was discussed in chapter 7. In this chapter I focus on the control of confounding.

Effect estimation

The basic effect measures, and methods of controlling confounding are described below. Usually, in epidemiology studies, we wish to measure the difference in disease occurrence between groups exposed and not exposed to a particular factor.

The analysis ideally should control simultaneously for all confounding factors. Control of confounding in the analysis involves stratifying the data according to the levels of the confounder(s) and calculating an effect estimate which summarizes the information across strata of the confounder(s). For example, controlling for age (grouped into 5 categories) and gender (with 2 categories) might involve grouping the data into the 10 (= 5 x 2) confounder strata and calculating a

summary effect estimate which is a weighted average of the stratum-specific effect estimates.

Confidence intervals

As well as estimating the effect of an exposure, it is also important to estimate the statistical precision of the effect estimate. The confidence interval (usually the 95% confidence interval) provides a range of values in which it is plausible (provided that there is no uncontrolled confounding or other bias) that the true effect estimate may lie. If the statistical model is correct, and there is no bias, then the confidence intervals derived from an infinite series of study repetitions would contain the true effect estimate with a frequency no less than its confidence level (Rothman and Greenland, 1998).

The usual practice is to use 90% or 95% confidence intervals, but these values are completely arbitrary. Given a large enough sample, an approximate 95% confidence interval for the true population mean is:

$$m \pm 1.96 \text{ SE}$$

where m is the observed mean of the sample, and SE is its standard error, estimated from the standard deviation of the sample divided by the square root of the sample size.

This confidence interval depends on two quantities (m and SE) which are estimated from the sample itself, and different results will be obtained from different samples. Provided that the samples are sufficiently large, then 95% of the time, the confidence interval estimated from the sample would contain the true population mean. One should note, however, that this is no guarantee that the interval from one's data contains the true value.

In most instances, epidemiologic data involves binomial (i.e. with persons in the denominator) or Poisson (i.e. with person-years in the denominator) outcome variables and ratio measures of effect. The estimated relative risk (rate ratio, risk ratio, odds ratio) has an approximate log normal distribution, and the $\ln(\text{RR})$ can be written as the difference of the two compared risks:

$$\ln(\text{RR}) = \ln(R_1/R_0) = \ln(R_1) - \ln(R_0)$$

Thus (assuming no bias) the 95% confidence interval for the natural log (\ln) of the relative risk is:

$$\ln(\text{RR}) \pm 1.96 \text{ SE}$$

Thus the confidence interval for the relative risk itself is:

$$\text{RR } e^{\pm 1.96 \text{ SE}}$$

P-Values

As discussed in chapter 5, the p-value is the probability that a test statistic as large or larger as that observed could have arisen by chance if there is no bias and if the null hypothesis (of no association between exposure and disease) is correct. The test statistic defines the p-value and usually has the form:

$$z = D/\text{SE}$$

where D is the observed difference and SE is the standard error of the difference.

This provides a test statistic (z) which can be used to calculate the probability (p-value) that a difference as large as that observed would have occurred by

chance if the null hypothesis (that there is no difference in reality) were true.

In the past, p-values have often been used to describe the results of a study as "significant" or "not significant" on the basis of decision rules involving an arbitrary alpha level as a "cutoff" for significance (e.g. $\alpha=0.05$). However, it is now recognised that there are major problems with this approach (Rothman and Greenland, 1998).

First, the p-value associated with a difference in outcome between two groups depends on two factors: the size of the difference; and the size of the study. A very small difference may be statistically significant if the study is very large, whereas a very large difference may not be significant if the study is very small. p-values thus combine two phenomena which should be kept separate: the size of the effect; and the size of the study used to measure it.

A second problem with significance testing is more fundamental. The purpose of significance testing is to reach a decision. However, in environmental research, decisions should ideally not be based on the results of a single study, but should be based on information from all available

studies, as well as non-statistical considerations such as the plausibility and coherence of the effect in the light of current theoretical and empirical knowledge (see chapter 13).

The problems of significance testing can be avoided by recognizing that the principal aim of an individual study should be to estimate the size of the effect rather than just to decide whether or not an effect is present. The point estimate should be accompanied by a confidence interval (the interval estimate) which indicates the precision of the point estimate by providing a range of values within which it is most plausible that the true treatment effect may lie if no bias were present (Gardner and Altman, 1986; Rothman and Greenland, 1998). The point estimate reflects the size of the effect, whereas the confidence interval reflects the study size on which this effect estimate is based. This approach also facilitates the comparison of the study findings with those of previous studies. Note that all conventional statistical methods assume "no bias is present". Because this assumption is rarely if ever correct, further considerations beyond the statistics presented here are always needed (see chapter 13).

12.2: Basic Analyses

Measures of Disease Occurrence

The basic measures of disease occurrence and association have been introduced in chapter 2. In this section I consider them in more depth and show how to calculate confidence intervals for the commonly

used measures. In the next section I extend these methods to adjust for potential confounders. I will only present "large sample" methods of analysis which have sample size requirements for valid

use. To avoid statistical bias, more complex techniques are required for analyses of studies involving very small numbers or sparse stratifications (Greenland et al, 2000). Once again, readers are referred to standard texts (particularly Rothman and Greenland, 1998) for a more comprehensive review of these methods. I will emphasise confidence intervals, but will also present methods for calculating p-values.

Table 12.1 shows the findings of a hypothetical incidence study of 20,000

persons followed for 10 years. As noted in chapter 2, three measures of disease incidence are commonly used in incidence studies.

The observed incidence rate in the non-exposed group (table 9.1) has the form:

$$I_0 = \frac{\text{cases}}{\text{person-time}} = \frac{b}{Y_0}$$

Table 12.1

Findings from a hypothetical cohort study of 20,000 persons followed for 10 years

	Exposed	Non-exposed	Ratio
Cases	1,813 (a)	952 (b)	
Non-cases	8,187 (c)	9,048 (d)	
Initial population size	10,000 (N ₁)	10,000 (N ₀)	
Person-years	90,635 (Y ₁)	95,163 (Y ₀)	
Incidence rate	0.0200 (I ₁)	0.0100 (I ₀)	2.00
Incidence proportion (average risk)	0.1813 (R ₁)	0.0952 (R ₀)	1.90
Incidence odds	0.2214 (O ₁)	0.1052 (O ₀)	2.11

The natural logarithm of I₀ has an approximate standard error (under the Poisson model for random variation in b) of:

$$SE [\ln(I_0)] = (1/b)^{0.5}$$

and an approximate 95% confidence interval for the incidence rate is thus:

$$I_0 e^{\pm 1.96 SE}$$

The observed incidence proportion in the non-exposed group has the form:

$$R_0 = \frac{\text{cases}}{\text{persons}} = \frac{b}{N_0}$$

The observed incidence proportion in the non-exposed group has the form:

$$R_0 = \frac{\text{cases}}{\text{persons}} = \frac{b}{N_0}$$

Its logarithm has an approximate standard error (under the binomial model for random variation in b) of:

$$SE[\ln(R_0)] = (1/b - 1/N_0)^{0.5}$$

and an approximate 95% confidence interval for the incidence proportion is thus:

$$R_0 e^{\pm 1.96 SE}$$

The observed incidence odds in the non-exposed group has the form:

$$O_0 = \frac{\text{cases}}{\text{non-cases}} = \frac{b}{d}$$

The natural log of the incidence odds ($\ln(O_0)$) has (under a binomial model) an approximate standard error of:

$$SE(\ln(O_0)) = (1/b + 1/d)^{0.5}$$

and a 95% confidence interval for O_0 is:

$$O_0 e^{\pm 1.96 SE}$$

These three measures of disease occurrence all involve the same numerator: the number of incident

cases of disease (b). They differ in whether their denominators represent person-years at risk (Y_0), persons at risk (N_0), or survivors (d).

Measures of Effect

Corresponding to these three measures of disease occurrence, there are three principal ratio measures of effect which can be used in incidence studies: the rate ratio, the risk ratio, and the odds ratio. In incidence case-control studies, the measure of effect is always the odds ratio (though what this is estimating depends on how the controls were chosen). In prevalence studies, the effect measure is usually the prevalence odds ratio, and the statistical methods are identical to those used in incidence case-control studies.

The observed incidence rate ratio has the form (table 12.1):

$$RR = \frac{I_1}{I_0} = \frac{a/Y_1}{b/Y_0}$$

An approximate p-value for the null hypothesis that the rate ratio equals the null value of 1.0 can be obtained using the person-time version of the Mantel-Haenszel chi-square (Breslow and Day, 1987). This test statistic compares the observed number of exposed cases with the number expected under the null hypothesis that $I_1 = I_0$:

$$\chi^2 = \frac{[\text{Obs}(a) - \text{Exp}(a)]^2}{\text{Var}(\text{Exp}(a))} = \frac{[a - Y_1 M_1 / T]^2}{[M_1 Y_1 Y_0 / T^2]}$$

where M_1 , Y_1 , Y_0 and T are as depicted in table 12.1.

The natural logarithm of the rate ratio has (under a Poisson model for a and b) an approximate standard error of:

$$SE[\ln(RR)] = (1/a + 1/b)^{0.5}$$

An approximate 95% confidence interval for the rate ratio is then given by (Rothman and Greenland, 1998):

$$RR e^{\pm 1.96 SE}$$

The risk ratio has the form:

$$RR = \frac{R_1}{R_0} = \frac{a/N_1}{b/N_0}$$

An approximate p-value for the null hypothesis that the risk ratio equals the null value of 1.0 can be obtained using the Mantel-Haenszel chi-square (Mantel and Haenszel, 1959):

$$\chi^2 = \frac{[\text{Obs}(a) - \text{Exp}(a)]^2}{\text{Var}(\text{Exp}(a))} = \frac{[a - N_1 M_1 / T]^2}{[M_1 M_0 N_1 N_0 / T^2 (T-1)]}$$

where M_1 , M_0 , N_1 , N_0 and T are as depicted in table 9.1.

The natural logarithm of the risk ratio has (under a binomial model for a and b) an approximate standard error of:

$$SE[\ln(RR)] = (1/a - 1/N_1 + 1/b - 1/N_0)^{0.5}$$

An approximate 95% confidence interval for the risk ratio is then given by:

$$RR e^{\pm 1.96 SE}$$

The incidence odds ratio has the form:

$$OR = \frac{O_1}{O_0} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

An approximate p-value for the hypothesis that the odds ratio equals the null value of 1.0 can be obtained from the Mantel-Haenszel chi-square (Mantel and Haenszel, 1959):

$$\chi^2 = \frac{[\text{Obs}(a) - \text{Exp}(a)]^2}{\text{Var}(\text{Exp}(a))} = \frac{[a - N_1 M_1 / T]^2}{[M_1 M_0 N_1 N_0 / T^2 (T-1)]}$$

where M_1 , M_0 , N_1 , N_0 and T are as depicted in table 9.1.

The natural logarithm of the odds ratio has (under a binomial model) an approximate standard error of:

$$SE[\ln(OR)] = (1/a + 1/b + 1/c + 1/d)^{0.5}$$

An approximate 95% confidence interval for the odds ratio is then given by:

$$OR e^{\pm 1.96 SE}$$

12.3: Control of Confounding

In general, control of confounding requires careful use of a priori knowledge, together with assessment of the extent to which the effect estimate changes when the factor is controlled in the analysis. Most epidemiologists prefer to make a decision based on the latter criterion, although it can be misleading, particularly if misclassification is present (Greenland and Robins, 1985a). The decision to control for a presumed confounder can certainly be made with more confidence if there is supporting prior knowledge that the factor is predictive of disease.

There are two methods of calculating a summary effect estimate to control confounding: pooling and standardisation (Rothman and Greenland, 1998).

Pooling

Pooling involves calculating a summary effect estimate assuming stratum-specific effects are equal. There are a number of different methods of obtaining pooled effect estimates, but a commonly used method which is both simple and close to being statistically optimal (even when there are small numbers in all strata) is the method of Mantel and Haenszel (1959).

The Mantel-Haenszel summary rate ratio has the form:

$$RR = \frac{\sum a_i Y_{0i} / T_i}{\sum b_i Y_{1i} / T_i}$$

where $T_i = Y_{1i} + Y_{0i}$

An approximate p-value for the null hypothesis that the summary rate ratio is 1.0 can be obtained from the person-time version of the one degree-of-freedom Mantel-Haenszel summary chi-square (Shore et al, 1976):

$$\chi^2 = \frac{[\sum \text{Obs}(a) - \sum \text{Exp}(a)]^2}{\sum \text{Var}(\text{Exp}(a))} = \frac{[\sum a_i - \sum Y_{1i} M_{1i} / T_i]^2}{[\sum M_{1i} Y_{1i} T_{0i} / T_i^2]}$$

where M_{1i} , Y_{1i} , Y_{0i} and T_i are as depicted in table 12.1.

An approximate standard error for the natural log of the rate ratio is (Greenland and Robins, 1985b):

$$SE = \frac{[\sum M_{1i} Y_{1i} Y_{0i} / T_i^2]^{0.5}}{[(\sum a_i Y_{0i} / T_i)(\sum b_i Y_{1i} / T_i)]^{0.5}}$$

Thus, an approximate 95% confidence interval for the summary rate ratio is then given by:

$$RR e^{\pm 1.96 SE}$$

The Mantel-Haenszel summary risk ratio has the form:

$$RR = \frac{\sum a_i N_{0i} / T_i}{\sum b_i N_{1i} / T_i}$$

An approximate p-value for the hypothesis that the summary risk ratio is 1.0 can be obtained from the one degree-of-freedom Mantel-Haenszel summary chi-square (Mantel and Haenszel, 1959):

$$\chi^2 = \frac{[\sum \text{Obs}(a) - \sum \text{Exp}(a)]^2}{\sum \text{Var}(\text{Exp}(a))} = \frac{[\sum a_i - \sum M_{1i}M_{1i}/T_i]^2}{[\sum M_{1i}M_{0i}M_{1i}N_{0i}/T_i^2(T_i-1)]}$$

where M_{1i} , M_{0i} , N_{1i} , N_{0i} and T_i are as depicted in table 9.1.

An approximate standard error for the natural log of the risk ratio is (Greenland and Robins, 1985b):

$$SE = \frac{[\sum M_{1i}N_{1i}N_{0i}/T_i^2 - \sum a_i b_i/T_i]^{0.5}}{[(\sum a_i N_{0i}/T_i)(\sum b_i N_{1i}/T_i)]^{0.5}}$$

Thus, an approximate 95% confidence interval for the summary risk ratio is then given by:

$$RR e^{\pm 1.96 SE}$$

The Mantel-Haenszel summary odds ratio has the form:

$$OR = \frac{\sum a_i d_i / T_i}{\sum b_i c_i / T_i}$$

An approximate p-value for the hypothesis that the summary odds ratio is 1.0 can be obtained from the one degree-of-freedom Mantel-Haenszel summary chi-square (Mantel and Haenszel, 1959):

$$\chi^2 = \frac{[\sum \text{Obs}(a) - \sum \text{Exp}(a)]^2}{\sum \text{Var}(\text{Exp}(a))} = \frac{[\sum a_i - \sum N_{1i}M_{1i}/T_i]^2}{[\sum M_{1i}M_{0i}N_{1i}N_{0i}/T_i^2(T_i-1)]}$$

where M_{1i} , M_{0i} , N_{1i} , N_{0i} and T_i are as depicted in table 12.1.

An approximate standard error for the natural log of the odds ratio (under a binomial or hypergeometric model) is (Robins et al, 1986):

$$SE = \frac{\sum PR}{2R_+^2} + \frac{\sum(PS + QR)}{2R_+S_+} + \frac{\sum QS}{2S_+^2}$$

where:

$$P = (a_i + d_i)/T_i$$

$$Q = (b_i + c_i)/T_i$$

$$R = a_i d_i / T_i$$

$$S = b_i c_i / T_i$$

$$R_+ = \sum R$$

$$S_+ = \sum S$$

Thus, an approximate 95% confidence interval for the summary odds ratio is then given by:

$$OR e^{\pm 1.96 SE}$$

Standardisation

Standardisation is an alternative approach to obtaining a summary effect estimate (Miettinen, 1974; Rothman and Greenland, 1998). Pooling involves calculating the effect estimate under the assumption that the measure (e.g. The rate ratio) would be the same (uniform) across strata if random error were absent. In contrast, standardisation involves taking a weighted average of the disease occurrence across strata (e.g. the standardized rate) and then comparing the standardized occurrence measure between exposed and non-exposed (e.g. the standardized rate ratio) with no assumptions of uniformity of effect. Standardisation is more prone than pooling to suffer from statistical instability due to small numbers in

specific strata; by comparison, pooling with Mantel-Haenszel estimators is robust and in general its statistical stability depends on the overall numbers rather than the numbers in specific strata. However, direct standardisation has practical advantages when more than two groups are being compared, e.g. when comparing multiple exposure groups or making comparisons between multiple countries or regions, and does not require the assumption of constant effects across strata.

The standardized rate has the form:

$$R = \frac{\sum w_i R_i}{\sum w_i}$$

The natural log of the standardized rate has an approximate standard error (under the Poisson model for random error) of:

$$SE = \frac{[\sum w_i^2 R_i / Y_i]^{0.5}}{R \sum w_i}$$

where Y_i is the person-time in stratum i . An approximate 95% confidence interval for the standardized rate is thus:

$$R e^{\pm 1.96 SE}$$

The standardized risk has the form:

$$R = \frac{\sum w_i R_i}{\sum w_i}$$

The natural log of the standardized risk has an approximate standard error

(under the binomial model for random error) of:

$$SE = \frac{[\sum w_i^2 R_i (1-R_i) / N_i]^{0.5}}{R \sum w_i}$$

where N_i is the number of persons in stratum i . An approximate 95% confidence interval for the standardized rate is thus:

$$R e^{\pm 1.96 SE}$$

Standardisation is not usually used for odds, since the odds is only used in the context of a case-control study, where the odds ratio is the effect measure of interest, but standardized odds ratios can be computed from case-control data (Miettinen, 1985; Rothman and Greenland, 1998).

A common choice of weights in international comparisons is Segi's World Population (Segi, 1960) shown in table 12.2, although it does reflect a "developed countries" bias in its age structure. In etiologic studies a better approach is to use the structure of the overall source population as the weights when calculating standardized rates or risks in subgroups of the source population. When one is specifically interested in the effects that exposure had, or would have, on a particular subpopulation, then weights should be taken from that subpopulation.

Multiple Regression

Multiple regression allows for the simultaneous control of more confounders by "smoothing" the data across confounder strata. In particular, rate ratios (based on person-time data) can be modelled using Poisson

log-linear rate regression, risk ratios can be modelled using binomial log-linear risk regression, and odds ratios can be modelled using binomial logistic regression (Pearce et al, 1988; Rothman and Greenland, 1998).

Similarly, continuous outcome variables (e.g. in a cross-sectional study) can be modelled with standard multiple linear regression methods. These models all have similar forms, with minor variations to take into account the different data types. They provide powerful tools when used appropriately, but are often used inappropriately, and should always be used in combination with the more straightforward methods presented here (Rothman and Greenland, 1998). Mathematical modelling methods and issues are reviewed in depth in a number of standard texts (e.g. Breslow and Day, 1980, 1987; Checkoway et al, 2004; Clayton and Hills, 1993; Rothman and Greenland, 1998), and will not be discussed in detail here.

Summary

The basic aim of the analysis of a single study is to estimate the effect of exposure on the outcome under study while controlling for confounding and minimizing other possible sources of bias. In addition, when confounding and other sources of bias cannot be removed, then it is important to assess their likely strength and direction. Control of confounding in the analysis involves stratifying the data according to the levels of the confounder(s) and calculating an effect estimate which summarizes the information across strata of the confounder(s). In general, control of confounding requires careful use of a priori knowledge, together with

Table 12.2

Segi's World population

Age-group	Population
0-4 years	12,000
5-9 years	10,000
10-14 years	9,000
15-19 years	9,000
20-24 years	8,000
25-29 years	8,000
30-34 years	6,000
35-39 years	6,000
40-44 years	6,000
45-49 years	6,000
50-54 years	5,000
55-59 years	4,000
60-64 years	4,000
65-69 years	3,000
70-74 years	2,000
75-79 years	1,000
80-84 years	500
85+ years	500
Total	100,000

Source: Segi (1960)

assessment of the extent to which the effect estimate changes when the factor is controlled in the analysis. There are two basic methods of calculating a summary effect estimate to control confounding: pooling and standardisation. Multiple regression allows for the simultaneous control of more confounders by "smoothing" the data across confounder strata. It provides a powerful tool when used appropriately, but are often used inappropriately, and should always be used in combination with the more straightforward methods presented here.

References

- Breslow NE, Day NE (1980). Statistical methods in cancer research. Vol I: The analysis of case-control studies. Lyon, France: IARC.
- Breslow NE, Day NE (1987). Statistical methods in cancer research. Vol II: The analysis of cohort studies. Lyon, France: IARC.
- Checkoway HA, Pearce N, Kriebel D (2004). Research methods in occupational epidemiology. 2nd ed. New York: Oxford University Press.
- Clayton D, Hills M (1993). Statistical models in epidemiology. Oxford: Oxford Scientific Publications.
- Dean J, Dean A, Burton A, Dicker R (1990). Epi Info. Version 5.01. Atlanta, GA: CDC.
- Gardner MJ, Altman DG (1986). Confidence intervals rather than p values: estimation rather than hypothesis testing. *Br Med J* 292: 746-50.
- Greenland S, Robins JM (1985a). Confounding and misclassification. *Am J Epidemiol* 122: 495-506.
- Greenland S, Robins JM (1985b). Estimation of a common effect parameter from sparse follow-up data. *Biometrics* 41: 55-68.
- Greenland S, Schwartzbaum JA, Finkle WD (2000). Problems due to small samples and sparse data in conditional logistic regression analysis. *Am J Epidemiol* 2000; 191: 530-9.
- Hills M, De Stavola BL (2002). A short introduction to Stata for biostatistics. London: Timberlake, 2002.
- Mantel N, Haenszel W (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 22: 719-48.
- Miettinen OS (1974). Standardization of risk ratios. *Am J Epidemiol* 96: 383-8.
- Miettinen OS (1985). Theoretical epidemiology. New York: Wiley and Sons.
- Pearce NE, Checkoway HA, Dement JM (1988). Exponential models for analyses of time-related factors: illustrated with asbestos textile worker mortality data. *J Occ Med* 30: 517-22.
- Robins JM, Breslow NE, Greenland S (1986). Estimation of the Mantel-Haenszel variance consistent with both sparse-data and large-strata limiting models. *Biometrics* 42: 311-23.
- Rothman KJ (2002). Epidemiology: an introduction. New York: Oxford University Press.
- Rothman KJ, Greenland S (1998). Modern epidemiology. 2nd ed. Philadelphia: Lippincott-Raven.
- Segi M (1960). Cancer mortality for selected sites in 24 countries (1950-1957). Sendai, Japan: Department of Public Health, Tohoku University School of Medicine.
- Shore RE, Pasternak BS, Curnen MG (1976). Relating influenza epidemics to childhood leukaemia in tumor registries without a defined population base. *Am J Epidemiol* 103: 527-35.

CHAPTER 13: Interpretation

[In: Pearce N. A Short Introduction to Epidemiology. 2nd ed. Wellington, CPHR, 2005]

In this chapter I first consider the issues involved in interpreting the findings of a single epidemiological study. I then consider problems of interpretation of all of the available evidence. Interpreting the findings of a single study includes considering the strength and precision of the effect estimate and the possibility that it may have been affected by various possible biases (confounding, selection bias, information bias). If it is concluded that the observed

associations are likely to be valid, then attention shifts to more general causal inference, which should be based on all available information. In both situations, it should be stressed that epidemiological studies almost always contain potential biases, and the focus should be on assessing the likely direction and magnitude of the biases, and whether they could explain the observed associations.

13.1: Appraisal of a Single Study

It is easy to criticize an epidemiological study. Populations do not usually randomize themselves by exposure status, do not always respond to requests to participate in epidemiological studies, may supply incomplete or inaccurate exposure histories for known or possible risk factors, and cannot be asked about unknown risk factors. Thus, although some studies are clearly better than others, it is important to emphasize that perfect epidemiological studies do not exist. Furthermore, it is usually not possible, nor desirable, to reach conclusions on the basis of the findings of a single study, and it is essential to consider all of the available evidence.

Nevertheless, when confronted with a new study, perhaps with unexpected findings, it is valuable to first consider possible explanations for the

associations found (or a lack of association) in the study, before proceeding to consider other evidence. However, the emphasis should not be on simply preparing a list of possible biases (e.g. Feinstein, 1988). Rather, it is essential to attempt to assess the likely strength and direction of each possible bias, and to assess whether these biases (and their possible interactions) could explain the observed associations.

What is the magnitude and precision of the effect estimate?

As discussed in chapter 5, random error (lack of precision) will occur in any epidemiologic study, just as it occurs in experimental studies. The possible role of random error is often addressed through the question "could the observed association be due to chance

alone?" and this issue is usually assessed by calculating the p-value. This is the probability (assuming that there are no biases) that a test statistic as large as that actually observed would be found in a study if the null hypothesis were true, i.e. that there was in reality no causal effect of exposure. However, recent reviews have stressed the limitations of p-values and significance testing (Rothman, 1978; Gardner and Altman, 1986; Poole, 1987; Pearce and Jackson, 1988). Foremost among these is that significance testing attempts to reach a decision on the basis of the data from a single study, whereas what is more important is the strength and precision of the effect estimate and whether the findings of a particular study are consistent with those of previous studies. These issues are better addressed by calculating confidence intervals rather than p-values (Gardner and Altman, 1986; Rothman and Greenland, 1998). Similarly, the possibility that the lack of a statistically significant association could be due to lack of precision (lack of study power) is more appropriately addressed by considering the confidence interval of the effect estimate rather than by making post hoc power calculations (Smith and Bates, 1992).

What are the likely strengths and directions of possible biases?

Systematic error is distinguished from random error in that it would be present even with an infinitely large study, whereas random error can be reduced by increasing the study size. Thus, systematic error, or "bias", occurs if there is a systematic difference between what the study is actually estimating and what it is intended to estimate. The types of bias (confounding, selection bias, information bias) have already been discussed in chapter 6. In the current context the key issue is that any

epidemiologic study will involve biases. The problem is not to identify possible biases (these will almost always exist), but rather to ascertain what direction they are likely to be in, and how strong they are is likely to be.

Confounding

In assessing whether an observed association could be due to confounding, the first consideration is whether all potential confounders have been appropriately controlled for or appropriately assessed (e.g. by collecting and using confounder information in a sample of study participants). If not, it is essential to assess the potential strength and direction of uncontrolled confounding.

In some areas of epidemiologic research, e.g. occupational and environmental studies, the strength of uncontrolled confounding is often less than might be expected. For example, Axelson (1978) has shown that for plausible estimates of the smoking prevalence in occupational populations, confounding by smoking can rarely account for a relative risk of lung cancer of greater than 1.5. Similarly, Siemiatycki et al (1988) have found that confounding by smoking is generally even weaker for internal comparisons in which exposed workers are compared with non-exposed workers in the same factory or industry). On the other hand, the potential for confounding can be severe in studies of lifestyle and related factors (e.g. diet, nutrition, exercise).

It is unreasonable to simply assume that a strong association could be due to confounding by unknown risk factors, since to be a strong confounder a factor must be a very strong risk factor as well as being strongly associated with exposure. For example, if an occupational study found a relative risk of 2.0 for lung cancer in exposed

workers, it is highly unlikely that this could be due to confounding by smoking, and it would be unreasonable to dismiss the study findings merely because smoking information had not been available. On the other hand, small relative risks (e.g. those in the range of 0.7-1.5, as frequently occur in dietary studies) are not so difficult to explain by lack of measurement, or poor measurement and control, of confounders.

Selection bias

Whereas confounding generally involves biases inherent in the source population, selection bias involves biases arising from the procedures by which the study subjects are chosen from the source population. As with confounding, if it is not possible to directly control for selection bias, it still may be possible to assess its likely strength and direction. It is unreasonable to dismiss the findings of a particular study because of possible selection bias, without at least attempting to assess which direction the possible selection bias would have been in, and how strong it might have been.

Information bias

With regards to information bias, the key issue is whether misclassification is likely to have been differential or non-differential. In the latter case, the bias will usually be in a known direction, i.e. towards the null. If misclassification has been differential, then it is important to attempt to assess what direction the bias is likely

to have been in. The important issue is not whether information bias could have occurred (this is almost always the case since there are almost always problems of misclassification of exposure and/or disease) but rather the likely direction and strength of such bias. In particular, if a study has yielded a positive finding (i.e. an effect estimate markedly different from the null value) then it is not valid to dismiss it because of the possibility of non-differential misclassification, or differential misclassification that is likely (although not guaranteed) to produce a bias towards the null.

Summary of Issues of Systematic Error

In summary, when assessing whether the findings of a particular study could be due to such biases, the important issue is not whether such biases are likely to have occurred (since they will almost always be present to some extent), but rather what their direction and strength is likely to be, and whether they taken together could explain the observed association. In particular, epidemiological studies are often criticized on the grounds that observed associations could be due to uncontrolled confounding or errors in the classification of exposure or disease. However, the likely strength of uncontrolled confounding is sometimes less than might be expected, and non-differential misclassification of exposure will usually (though not always) produce a tendency for false negative findings rather than false positive findings.

13.2: Appraisal of All of the Available Evidence

If it is concluded that the association in a particular study is unlikely to be primarily due to bias and chance, attention then shifts to assessing whether this association exists more generally, and whether the association is likely to be causal. This should involve a review of all of the available evidence including non-epidemiological studies. A systematic quantitative review of the epidemiological evidence may involve a formal meta-analysis with statistical pooling of information from the various studies (e.g. Dickerson and Berlin, 1992; Rothman and Greenland, 1998). However, such a summary of the various study findings is just one step in the process of causal inference. A systematic approach to causal inference was elaborated by Hill (1965) and has since been widely used and adapted (e.g. Beaglehole et al (1993)). I will divide these considerations into those that involve systematic review of the epidemiological evidence (including meta-analyses) and those that also involve consideration of evidence from animal or mechanistic studies.

Evidence From Epidemiological Studies

Considerations for assessing the epidemiological evidence include temporality, specificity, consistency, strength of association and whether there is evidence of a dose-response relationship (Hill, 1965).

Temporality is crucial; the cause must precede the effect. This is usually self-evident, but difficulties may arise in studies (particularly case-control studies) when measurements of exposure and effect are made at the

same time (e.g. by questionnaire, or blood tests, etc).

The criterion of specificity has been criticised (e.g. Rothman and Greenland, 1998), on the grounds that there are many instances of exposures that have multiple (i.e. non-specific) effects. These include tobacco smoke and ionizing radiation, both of which cause many different types of cancer. Nevertheless, the specificity of the effect may be relevant in assessing the possibility of various biases. For example, if an exposure is associated with esophageal cancer but is not associated with lung cancer, then the association is unlikely to be due to confounding by smoking.

Consistency is demonstrated by several studies giving similar results, and corresponds to the statistical concept of homogeneity across studies (Rothman and Greenland, 1998). This is particularly important when a variety of designs are used in different settings, since the likelihood that all studies are all suffering from the same biases may thereby be reduced. On the other hand, a lack of consistency does not exclude a causal association, because different exposure levels and other conditions may alter the effect of exposure in certain studies.

The strength of association is important in that a relative risk that is far from the null value of 1.0 is more likely to be causal than a weak association, which could be more easily explained by confounding or other biases. However, the fact that an association is weak does not preclude it from being causal; rather it means that it is more difficult to

exclude alternative explanations for the observed association.

A dose-response relationship occurs when changes in the level of exposure are associated with changes in the prevalence or incidence of the effect than one would expect from biologic considerations. The absence of an expected dose-response relationship provides evidence against a causal relationship, while the presence of an expected relationship narrows the scope of biases that could explain the relationship.

Experimental evidence provides strong evidence of causality, but this is rarely available for occupational exposures.

Meta-Analysis

In the past, epidemiological evidence has been assessed in literature reviews, but in recent years there has been an increasing emphasis on formal meta-analysis, i.e. systematic quantitative reviews. One benefit of a meta-analysis is that it can reduce the probability of false negative results because of small numbers in specific studies (Egger and Davey-Smith, 1997), and may enable the effect of an exposure to be estimated with greater precision than is possible in a single study. Furthermore, although a meta-analysis should ideally be based on individual data, relatively simple methods are available for meta-analyses of published studies in which the study (rather than the individual) is the unit of statistical analysis (Rothman and Greenland, 1998). Such methods can be used to address the causal considerations outlined above, in particular the overall strength of association and the shape and strength of the dose-response curve. Just as importantly, statistical methods can also be used to assess consistency between studies, but because statistical tests for

homogeneity often have relatively low power, it is more appropriate to examine the magnitude of variation instead of relying on formal statistical tests (Rothman and Greenland, 1998).

The limitations of meta-analyses should also be emphasized (Greenland, 1994; Egger and Davey-Smith, 1997; Egger et al, 1997). Strikingly different results can be obtained depending on which studies are included in a meta-analysis. Publication bias is of particular concern, given the tendency of journals to publish "positive findings" and for the publication of "negative findings" to be delayed (Egger and Davey-Smith, 1998), but naive graphical approaches to its assessment can be misleading (Greenland, 1994).

Even when an "unbiased" and comprehensive list of studies is included in a meta-analysis, there still remain the same problems of selection bias, information bias, and confounding, that need to be addressed in assessing individual studies. Thus, a systematic quantitative review (i.e. meta-analysis) is like a report of a single study in that both quantitative and narrative elements are required to produce a balanced picture (Rothman and Greenland, 1998). Essentially the same issues need to be addressed as in a report of a single study: what is the overall magnitude and precision of the effect estimate (if it is considered appropriate to calculate a summary effect estimate), and what are likely strengths and directions of possible biases?

An advantage of meta-analysis is that these issues can often be better addressed by contrasting the findings of studies based on different populations, or using different study designs. Thus, possible systematic biases can be addressed with actual data from specific studies rather than by hypothetical

examples. For example, in a study of an occupational exposure and lung cancer, there might be concern that an observed association was due to confounding by smoking. If smoking data had not been available, then the best that could be done would be to attempt to assess the likely extent of confounding by smoking (see chapter 6), for example by sensitivity analysis (Rothman and Greenland, 1998). However, in a meta-analysis, if smoking information were available for some (but not all) studies then these studies could be examined to assess the likely strength and direction of confounding by smoking (if any).

Similarly, studies of exposure to phenoxy herbicides and the development of soft tissue sarcoma and non-Hodgkin's lymphoma have produced widely differing findings, and it has been suggested that the high relative risks obtained in the Swedish studies could be due to "recall bias" (a particular type of information bias) in that cases or cancer (soft tissue sarcoma or non-Hodgkin's lymphoma) were compared with healthy general population controls, and that patients with cancer may be more likely to recall previous chemical exposures. This hypothesis was tested in specific studies (e.g. Hardell et al, 1979, 1981), but can also be tested more generally by considering the findings of studies that used general population controls with those that used "other cancer" controls. In particular, one New Zealand study (Pearce et al, 1986) used both types of controls and found similar results with each, indicating that recall bias was not an important problem in this study.

In summary, a key advantage of meta-analysis is that pooling findings from studies will increase the numbers

available for analysis and will therefore reduce random error. However, it will not necessarily reduce systematic error, and may even increase it (because of publication bias). Nevertheless, a careful meta-analysis will enable various possible biases to be addressed, using actual data from specific studies, rather than hypothetical examples. Such a meta-analysis will therefore facilitate the consideration of the causal considerations listed above, and in some instances will provide a valid summary estimate of the overall strength of association and the shape and strength of the dose-response curve (Greenland, 2003).

Combination of Epidemiological Evidence With Evidence From Other Sources

Epidemiological evidence should be considered together with all other available evidence, including animal experiments. An association is plausible if it is consistent with other knowledge, whereas the epidemiological evidence is coherent if it is not inconsistent with other knowledge. For instance, laboratory experiments may have shown that a particular environmental exposure can cause cancer in laboratory animals, and this would make more plausible the hypothesis that this exposure could cause cancer in humans. However, biological plausibility is a relative concept; many epidemiological associations were considered implausible when they were first discovered but were subsequently confirmed by other evidence, e.g. the relation of lice to typhus. Lack of plausibility may simply reflect lack of knowledge (medical, biological, or social) which is continually changing and evolving.

Summary

The task of interpreting the findings of a single epidemiological study should be differentiated from that of interpreting all of the available evidence. Interpreting the findings of a single study includes considering the strength and precision of the effect estimate and the possibility that it may have been affected by various possible biases (confounding, selection bias, information bias). The important issue is not whether such biases are likely to have occurred (since they will almost always be present to some extent), but rather what their direction and strength is likely to be, and whether together they could explain the observed association. If the observed associations seem likely to be valid, then attention shifts to more general causal inference, which should be based on all available information. This includes assessing the specificity, strength and consistency of the association and the dose-response across all epidemiological studies. This may include the use of meta-analysis, but it is often not appropriate to derive a single summary effect estimate across all studies. Rather, a meta-analysis can be used to examine hypotheses about reasons for

differences between study findings and the likely magnitude of possible biases. Furthermore, causal inference also necessitates considering non-epidemiological evidence from other sources (animal studies, mechanistic studies) in the consideration of more general causal criteria including the plausibility and coherence of the overall evidence.

Despite the continual need to assess possible biases, and to consider possible imperfections in the epidemiological data, it is also important to ensure that preventive action occurs when this is warranted, albeit on the basis of imperfect data. As Hill (1965) writes:

"All scientific work is incomplete - whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us a freedom to ignore the knowledge that we already have, or to postpone the action that it appears to demand at a given time."

References

Axelsson O (1978). Aspects on confounding in occupational health

epidemiology. Scand J Work Environ Health 4: 85-9.

- Beaglehole R, Bonita R, Kjellstrom T (1993). Basic epidemiology. Geneva: WHO.
- Dickerson K, Berlin JA (1992). Meta-analysis: state-of-the-science. *Epidemiologic Reviews* 14: 154-76.
- Egger M, Davey-Smith G (1997). Meta-analysis: principles and promise. *Br Med J* 1997; 315: 1371-4.
- Egger M, Davey-Smith G, Phillips A (1997). Meta-analysis: principles and procedures. *Br Med J* 1997; 315: 1533-7.
- Egger M, Davey-Smith G (1998). Meta analysis: bias in location and selection of studies. *Br Med J* 1998; 316: 61-6.
- Feinstein AR (1988). Scientific standards in epidemiologic studies of the menace of daily life. *Science* 242: 1257-63.
- Gardner MJ, Altman DG (1986). Confidence intervals rather than p values: estimation rather than hypothesis testing. *Br Med J* 292: 746-50.
- Greenland S (1994). A critical look at some popular meta-analytic methods. *Am J Epidemiol* 140: 290-6.
- Greenland S (2003). The impact of prior distributions for uncontrolled confounding and response bias: a case study of the relation of wire codes and magnetic fields to childhood leukemia. *J Am Statist Assoc* 98: 1-8.
- Hardell L, Sandstrom A (1979). Case-control study: soft-tissue sarcomas and exposure to phenoxyacetic acids or chlorophenols. *Br J Cancer* 39: 711-7.
- Hardell L, Eriksson M, Lenner P, Lundgren E (1981). Malignant lymphoma and exposure to chemicals, especially organic solvents, chlorophenols and phenoxy acids: a case-control study. *Br J Cancer* 43: 169-76.
- Hill AB (1965). The environment and disease: association of causation? *Proc R Soc Med* 58: 295-300.
- Pearce NE, Smith AH, Howard JK, et al (1986). Non-Hodgkin's lymphoma and exposure to phenoxyherbicides, chlorophenols, fencing work and meat works employment: a case-control study. *Brit J Ind Med* 43: 75-83.
- Pearce NE, Jackson RT (1988). Statistical testing and estimation in medical research. *NZ Med J* 101: 569-70.
- Poole C (1987). Beyond the confidence interval. *AJPH* 77: 195-9.
- Rothman KJ (1978). A show of confidence. *N Engl J Med* 299: 1362-3.
- Rothman KJ, Greenland S (1998). *Modern epidemiology*. 2nd ed. Philadelphia: Lippincott-Raven.
- Siemiatycki J, Wacholder S, Dewar R, et al (1988). Smoking and degree of occupational exposure: Are internal analyses in cohort studies likely to be confounded by smoking status? *American Journal of Industrial Medicine* 13:59-69.
- Smith AH, Bates M (1992). Confidence limit analyses should replace power calculations in the interpretation of epidemiologic studies. *Epidemiol* 3: 449-52.